Diary No
Date 1-2.
Lib
Bureau Edu... Research

Journal of Experimental Psychology

DAVID A. GRANT, Editor
University of Wisconsin—Madison

HAROLD W. HAKE, Associate Editor
University of Illinois, Urbana

LEO POSTMAN, Associate Editor
University of California, Berkeley

LEONARD E. ROSS, Associate Editor
University of Wisconsin—Madison

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

NANCY S. ANDERSON, *University of Maryland*
NORMAN H. ANDERSON, *University of California, San Diego*
WILLIAM H. BATCHELDER, *University of California, Irvine*
WILLIAM F. BATTIG, *University of Colorado*
LYLE E. BOURNE, JR., *University of Colorado*
GEORGE E. BRIGGS, *Ohio State University*
E. J. CAPALDI, *Purdue University*
VIRGIL R. CARLSON, *National Institute of Mental Health*
HERBERT H. CLARK, *Stanford University*
ANDREW CRIDER, *Williams College*
HELEN B. DALY, *State University of New York, College at Oswego*
ROGER L. DOMINOWSKI, *University of Illinois at Chicago Circle*
SHELDON M. EBENHOLTZ, *University of Wisconsin—Madison*
WARD EDWARDS, *University of Michigan*
HOWARD EGETH, *Johns Hopkins University*
BRUCE R. EKSTRAND, *University of Colorado*
HENRY C. ELLIS, *University of New Mexico*
JAMES R. ERICKSON, *Ohio State University*
MURRAY GLANZER, *New York University*
JUDITH GOGGIN, *University of Texas at El Paso*
FRANCES K. GRAHAM, *University of Wisconsin—Madison*
WILLIAM W. GRINGS, *University of Southern California*
ROBERT C. HAYGOOD, *Arizona State University*
ELIOT HEARST, *Indiana University*
JULIAN HOCHBERG, *Columbia University*
LEONARD M. HOROWITZ, *Stanford University*
WILLIAM C. HOWELL, *Rice University*
STEWART H. HULSE, *Johns Hopkins University*
GEOFFREY KEPPEL, *University of California, Berkeley*
GREGORY A. KIMBLE, *University of Colorado*
HERBERT D. KIMMEL, *University of South Florida*
PATRICK R. LAUGHLIN, University of Illinois
HERSCHEL W. LEIBOWITZ, *Pennsylvania State University*
FRANK A. LOGAN, *University of New Mexico*
IRVING MALTZMAN, *University of California, Los Angeles*
DONALD R. MEYER, *Ohio State University*
BENNET B. MURDOCK, JR., *University of Toronto*
RAYMOND S. NICKERSON, *Bolt, Beranek, and Newman, Inc., Cambridge, Massachusetts*
MERRILL E. NOBLE, *Pennsylvania State University*
J. BRUCE OVERMIER, *University of Minnesota*
MICHAEL I. POSNER, *University of Oregon*
WILLIAM F. PROKASY, *University of Utah*
L. STARLING REID, *University of Virginia*
DONALD A. RILEY, *University of California, Berkeley*
WILLARD N. RUNQUIST, *University of Alberta*
RUDOLPH SCHULZ, *University of Iowa*
BRYAN E. SHEPP, *Brown University*
J. RICHARD SIMON, *University of Iowa*
PAUL SLOVIC, *Oregon Research Institute*
JANET T. SPENCE, *University of Texas at Austin*
HERBERT H. STENSON, *University of Illinois at Chicago Circle*
DAVID R. THOMAS, *University of Colorado*
MILTON A. TRAPOLD, *University of Minnesota*
ENDEL TULVING, *Yale University*
WILLIAM P. WALLACE, *University of Nevada*
WAYNE A. WICKELGREN, *University of Oregon*
GORDON WOOD, *Michigan State University*

HAROLD P. VAN COTT
Managing Editor

RITA JASKOWITZ
BETH FORRESTER
Technical Editors

HENRY S. SALMON
Advertising and Business Manager

ANNE REDMAN
Subscription Manager

Bureau ... Research
6.3.81
J 709
$5.00

Volume 96, 1972

PUBLISHED MONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA. 17604
AND 1200 SEVENTEENTH ST. N. W., WASHINGTON, D. C. 20036

Second-class postage paid at Lancaster, Pa.

CONTENTS OF VOLUME 96

Angel, R. W., and Garland, H. Transfer of Information from Manual to Oculomotor Control System ... 92
Atkinson, R. C. Optimizing the Learning of a Second-Language Vocabulary ... 124
Atkinson, R. C. See Darley, C. F.
Baboorian, N. See Checkosky, S. F.
Bacharach, V. R. Semantic and Syntactic Constraints on Free-Recall Learning of Sentential Material ... 223
Bainbridge, L., and Sanders, M. The Generality of Fitts's Law ... 130
Barrett, T. R., and Ekstrand, B. R. Effect of Sleep on Memory: III. Controlling for Time-of-Day Effects ... 321
Barton, A. K. Picture versus Word and Relevant Value "Relatedness" in Rule-Learning Problems ... 208
Bartz, W. H., Lewis, M. Q., and Swinton, G. Serial Position Effects for Repeated Free Recall: Negative Recency or Positive Primacy? ... 10
Beaumont, G. See Dimond, S.
Beito, A. See Wallace, W. P.
Berlyne, D. E., and Normore, L. F. Effects of Prior Uncertainty on Incidental Free Recall 43
Brace, T. G. See Suboski, M. D.
Broerse, J. See Over, R.
Brown, E. L. See Miscik, J. G.
Capaldi, E. J. Successive Negative Contrast Effect: Intertrial Interval, Type of Shift, and Four Sources of Generalization Decrement ... 433
Cermak, L. S., and Sampson, J. B. Decay of Acoustic Proactive Facilitation ... 237
Chase, W. G. See Chi, M. T. H.
Checkosky, S. F., and Baboorian, N. Memory Search for CVC and CCC Trigrams ... 158
Chenoweth, R. E. See Laughlin, P. R.
Chi, M. T. H., and Chase, W. G. Effects of Modality and Similarity on Context Recall ... 219
Cohn, N. H. See Sgro, J. A.
Coles, M. G. H. Cardiac and Respiratory Activity during Visual Search ... 371
Collins, J. K. See Singer, G.
Connor, J. M. Serial and Parallel Encoding Processes in Memory and Visual Search ... 363
Crassini, B. See Over, R.
Daniel, T. C. Nature of the Effect of Verbal Labels on Recognition Memory for Form ... 152
Darley, C. F., Klatzky, R. L., and Atkinson, R. C. Effects of Memory Load on Reaction Time ... 232
Davenport, J. W. See Flaherty, C. F.
Davenport, W. G. See Roy, E. A.
Davis, J. C., and Smith, M. C. Memory for Unattended Input ... 380
Deffenbacher, K. A. See Miscik, J. G.
DeRosa, D. V. See Kaminsky, C. A.
Dick, A. O. Parallel and Serial Processing in Tachistoscopic Recognition: Two Mechanisms ... 60
Dieter, R. See Suboski, M. D.
Dimond, S., and Beaumont, G. Hemisphere Function and Color Naming ... 87
Drake, R. W., Jr., and Meyer, P. A. Immediate, Generalized, and Enduring Effectiveness of Punishment and Response Prevention of Human Avoidance Responding ... 110
Dudley, S. D. See Sgro, J. A.
Duffey, T. M. See McMurray, D. W.
Dyck, D. G. See Lamberth, J.
Edwards, W. See Snapper, K. J.
Eggemeier, F. T. See Wickens, D. D.
Ekstrand, B. R. See Barrett, T. R.
Farrell, B. B. See Laughlin, P. R.
Faust-Adams, A. S. Interference in Short-Term Retention of Discrete Movements ... 400
Flaherty, C. F., and Davenport, J. W. Successive Brightness Discrimination in Rats following Regular versus Random Intermittent Reinforcement ... 1
Freundl, P. C., and Senf, G. M. Effect of Stimulus Rate, Material, and Storage Instructions on Recall of Bisensory Items: Storage or Retrieval Effects? ... 338
Fritzen, J., and Johnson, N. F. Functional Units in Free Recall ... 226
Furedy, J. J. See Schiffman, K.
Garland, H. See Angel, R. W.
Geller, E. S., and Whitman, C. P. Prediction Outcome and Choice Reaction Time: A Memory-Dependent Relationship ... 334
Geller, E. S. See Whitman, C. P.
Glassman, W. E. Subvocal Activity and Acoustic Confusions in Short-Term Memory ... 164
Greenstein, G. J. See Houck, R. L.
Griffith, D. See Johnston, W. A.
Hamm, N. H. See Miscik, J. G.
Hammond, O. W. See Staats, A. W.
Hamon, T. G. See Young, R. K.
Hansen, G. See Thomas, D. R.
Hawkins, H. L., and Shigley, R. H. Irrelevant Information and Processing Mode in Speeded Discrimination ... 389
Helson, H. See Kitterle, F. L.
Herrmann, D. J. See McLaughlin, J. P.
Hershenson, M. Verbal Report and Visual Matching Latency as a Function of the Pronounceability of Letter Arrays ... 104
Hicks, R. E., and Young, R. K. Part–Whole List Transfer in Free Recall ... 328
Houck, R. L., Mefferd, R. B., Jr., and Greenstein, G. J. Influence of a Visual Frame and Vertical–Horizontal Illusion on Shape and Size Perception ... 273

Izawa, C. Retention Interval Hypothesis and Evidence for Its Basic Assumptions 17
Jahnke, J. C. Distribution of Repeated and Nonrepeated Target Elements and Short-Term Recognition Memory 345
Jarrold, L. A. See Suboski, M. D.
Johnson, N. F. See Fritzen, J.
Johnston, W. A., Wagstaff, R. R., and Griffith, D. Information-Processing Analysis of Verbal Learning 307
Kaminsky, C. A., and DeRosa, D. V. Influence of Retrieval Cues and Set Organization on Short-Term Recognition Memory 449
Karsh, R., and Monty, R. A. Irrelevant Information, Irregularity, and the Pacing of Rehearsal in Sequential Short-Term Memory 142
Kelley, J. See Weber, R. J.
Kirby, N. H. Sequential Effects in Serial Reaction Time 32
Kitterle, F. L., and Helson, H. On the Inhibitory Effect of a Second Stimulus following the Primary Stimulus to React 138
Kjellberg, A. See Lisper, H.-O.
Klatzky, R. L. Visual and Verbal Coding of Laterally Presented Pictures 439
Klatzky, R. L. See Darley, C. F.
Lamberth, J., and Dyck, D. G. Reward Magnitude and Sequence of Magnitudes as Determinants of Resistance to Extinction in Humans 280
Laughlin, P. R., Chenoweth, R. E., Farrell, B. B., and McGrath, J. E. Concept Attainment as a Function of Motivation and Task Complexity 54
Levin, I. P. See Norman, K. L.
Lewis, J. L. Semantic Processing with Bisensory Stimulation 455
Lewis, M. Q. See Bartz, W. H.
Libb, J. W. Frustration and Task Complexity 67
Light, L. L. Homonyms and Synonyms as Retrieval Cues 255
Lisper, H.-O., and Kjellberg, A. Effects of 24-Hour Sleep Deprivation on Rate of Decrement in a 10-Minute Auditory Time Task 287
Little, S. See Weber, R. J.
Lively, B. L. Speed/Accuracy Trade Off and Practice as Determinants of Stage Durations in a Memory-Search Task 97
Lodwig, A. K. Experimental Manipulation of Orienting Reflex during Semantic Conditioning 416
Loftus, E. F. Nouns, Adjectives, and Semantic Memory 213
Luster, T. Vestibular Adaptation in Man: Effects of Increased Acceleration during Different Phases of Adaptation 263
McGrath, J. E. See Laughlin, P. R.
McLaughlin, J. P., and Herrmann, D. J. Recognition of a Categorized List 235
McMurray, D. W., and Duffy, T. M. Meaningfulness and Pronounceability as Chunking Units in Short-Term Memory 291
Mefferd, R. B., Jr. See Houck, R. L.
Meyer, P. A. See Drake, R. W., Jr.
Millar, S. Effects of Interpolated Tasks on Latency and Accuracy of Intramodal and Cross-Modal Shape Recognition by Children 170
Miller, E. A. Interaction of Vision and Touch in Conflict and Nonconflict Form Perception Tasks 114
Miller, J. T. See Thomas, D. R.
Miscik, J. G., Smith, J. M., Hamm, N. H., Deffenbacher, K. A., and Brown, E. L. Short-Term Retention of Auditory Sequences as a Function of Stimulus Duration, Interstimulus Interval, and Encoding Technique 147
Monty, R. A. See Karsh, R.
Newby, R. W., and Young, R. K. Percentage of Occurrence of Correct Response and Implicit Associative Responses in Verbal Discrimination Learning 49
Newby, R. W., and Young, R. K. Presentation of Correct and Incorrect Items in Verbal Discrimination Learning 471
Newby, R. W. See Young, R. K.
Newman, C. V. Familiar and Relative Size Cues and Surface Texture as Determinants of Relative Distance Judgments 37
Norman, K. L., and Levin, I. P. Tests of an All-or-None Model of Verbal Mediated Responding 247
Normore, L. F. See Berlyne, D. E.
Over, R., Broerse, J., and Crassini, B. Orientation Illusion and Masking in Central and Peripheral Vision 25
Pavlik, W. B. See Sgro, J. A.
Peterson, C. R. See Snapper, K. J.
Posnansky, C. J. Probing for the Functional Stimuli in Serial Learning 184
Postman, L., and Warren, L. Test of the Total-Time Hypothesis in Free-Recall Learning .. 176
Pschirrer, M. Goal Events as Discriminative Stimuli over Extended Intertrial Intervals .. 425
Remington, R. K. See Wallace, W. P.
Reutener, D. B. See Wickens, D. D.
Roy, E. A., and Davenport, W. G. Factors in Motor Short-Term Memory: The Interference Effect of Interpolated Activity 134
Russell, D. G. See Schmidt, R. A.
Sampson, J. B. See Cermak, L. S.
Sanders, M. See Bainbridge, L.
Schiffman, K., and Furedy, J. J. Failures of Contingency and Cognitive Factors to Affect Long-Interval Differential Pavlovian Autonomic Conditioning 215
Schmidt, R. A., and Russell, D. G. Movement Velocity and Movement Time as Determiners of Degree of Preprogramming in Simple Movements 315
Seamon, J. G. Imagery Codes and Human Information Retrieval 468

Seamon, J. G. Serial Position Effects in Probe Recall: Effect of Rehearsal on Reaction Time 460
Senf, G. M. See Freundl, P. C.
Sgro, J. A., Cohn, N. H., and Dudley, S. D. Partial Delay of Reward in the Double Alleyway 458
Sgro, J. A., Pavlik, W. B., Showalter, J. R., and Cohn, N. H. Effects of Partial Reinforcement in One or Both Goal Boxes of a Double Alleyway 229
Shigley, R. H. See Hawkins, H. L.
Showalter, J. R. See Sgro, J. A.
Simpson, P. J. High-Speed Memory Scanning: Stability and Generality 239
Singer, G., and Collins, J. K. Dissipation of Sensory Aftereffects as a Function of Axon Length 194
Smith, J. M. See Miscik, J. G.
Smith, M. C. See Davis, J. C.
Snapper, K. J., Edwards, W., and Peterson, C. R. Information Purchase Strategies of Gamblers and Students 211
Staats, A. W., and Hammond, O. W. Natural Words as Physiological Conditioned Stimuli: Food-Word-Elicited Salivation and Deprivation Effects 206
Suboski, M. D., Brace, T. G., Jarrold, L. A., Teller, K. J., and Dieter, R. Interstimulus Interval and Time Estimation in Ratings of Signaled Shock Aversiveness 407
Swinton, G. See Bartz, W. H.
Taylor, R. L. Reading Spatially Transformed Digits 396
Teller, K. J. See Suboski, M. D.
Ternes, W., and Yuille, J. C. Words and Pictures in an STM Task 78
Thomas, D. R., Miller, J. T., and Hansen, G. Role of Stimulus Comparison in Equivalence Training 297
Ullrich, J. R. Re-pairing, Number of Alternatives, and Meaningfulness in Verbal Discrimination Learning 201
Wade, N. J. Effect of Forward Head Inclination on Visual Orientation during Lateral Body Tilt 203
Wagstaff, R. R. See Johnston, W. A.
Wallace, W. P., Remington, R. K., and Beito, A. Retroactive Inhibition as a Function of Transfer Paradigm in Verbal Discrimination 463
Warren, L. See Postman, L.
Wasilewski, B. K. Effect of Accuracy-Emphasized Instructions on Performance on an Attribute–Identification Task 199
Weber, R. J., Kelley, J., and Little, S. Is Visual Image Sequencing under Verbal Control? 354
Whitman, C. P., and Geller, E. S. Effects of Stimulus and Response Patterns on Choice Reaction Time 466
Whitman, C. P. See Geller, E. S.
Wickens, D. D., Reutener, D. B., and Eggemeier, F. T. Sense Impression as an Encoding Dimension of Words 301
Williams, J. D. Effects of Practice with Controlled Stimulus Pairs on Same–Different Judgments 73
Young, R. K., Newby, R. W., and Hamon, T. G. Acquisition and Familiarization Hypotheses in Paired-Associate Learning 473
Young, R. K. See Hicks, R. E.
Young, R. K. See Newby, R. W.
Yuille, J. C. See Ternes, W.

Diary No [illegible]
Date 1.2
il. N Lib.
Bureau Ed.l Psy. Research.

Journal of
Experimental Psychology

Vol. 96, No. 1 November 1972

SUCCESSIVE BRIGHTNESS DISCRIMINATION IN RATS FOLLOWING REGULAR VERSUS RANDOM INTERMITTENT REINFORCEMENT[1]

CHARLES F. FLAHERTY[2]
Rutgers—The State University

JOHN W. DAVENPORT
University of Wisconsin

Two groups of rats were given 50% random or regular reinforcement in a gray runway until the regular-trained group displayed clear evidence of patterning. In a subsequent successive discrimination in black and white runways, the performance of the pattern-trained group was superior to that of the random-trained group, particularly with respect to terminal speeds on S− trials. The discrimination results confirmed A. Amsel's notions regarding the mediated transfer of persistence and avoidance by frustration. The within-session development of patterning behavior in the gray runway task was found to be consistent with B. B. Murdock's 1960 scale of stimulus distinctiveness for serial learning situations. E. J. Capaldi's sequential theory is discussed in relation to both the patterning and discrimination results.

Rats develop differential response tendencies on rewarded (R) and nonrewarded (N) trials in a 50% reinforcement schedule if the R and N are regularly alternated (e.g., Capaldi & Stanley, 1963; Tyler, Wortz, & Bitterman, 1953). Terminal performance, called patterning, on such a schedule consists of fast responding on R trials and slow responding on N trials, in the absence of external discriminative cues. Capaldi (1967) has attributed these differential response tendencies to the conditioning of stimulus aftereffects, S^R and S^N, of the R and N trials, respectively. Thus S^N is conditioned to the approach response because an N trial is always followed by an R trial, and conversely, S^R acquires an inhibitory influence because an R trial is always followed by an N trial. Commonly used random 50% schedules, on the other hand, do not generate patterning because neither S^R nor S^N is consistently associated with reward (or nonreward) over many trials, with the result that these stimulus aftereffects provide little or no discriminative control over responding.

Although there is now a considerable literature on the blocking or enhancement of stimulus control as a function of prior training on another dimension (e.g., Mackintosh & Honig, 1970; Thomas, Freeman, Svinicki, Burr, & Lyons, 1970) and intensive work reported (e.g., Amsel & Ward, 1965) on black–white runway discrimination as a function of prediscrimination experience with the black and white discriminanda, the question of how perform-

---

[1] This work was supported by a grant from the Rutgers Research Council and a Biological Sciences support grant to the first author and by Grant RR-00167 from the National Institutes of Health to the Wisconsin Regional Primate Research Center, Publication 11-053 of the Wisconsin Regional Primate Research Center.

[2] Requests for reprints should be sent to Charles F. Flaherty, Department of Psychology, Rutgers—The State University, New Brunswick, New Jersey 08903.

© 1972 by the American Psychological Association, Inc.

ance in an externally cued discrimination is affected by prior learning of a discrimination which does not involve external discriminative cues has apparently been neglected. It has occurred to us that the more specific question of how the acquisition of a successive black–white discrimination in dual runways proceeds in rats which have previously acquired patterning on a regular 50% schedule, as compared with rats having prior exposure to a random 50% schedule, in a single gray runway is of particular theoretical interest.

Amsel's (1967) theoretical conceptions about anticipatory frustration (r_F), frustration-produced stimuli (s_F), and various s_F-elicited response tendencies (e.g., persisting approach, avoidance, or competing responses) which can transfer from one learning situation to another via the mediating $r_F \rightarrow s_F$ mechanism lead us to expect that rats who learn patterning in a gray alley would acquire a subsequent discrimination in black and white alleys more rapidly than would rats receiving random 50% reinforcement in the gray alley, and would display the discrimination in terms of greater speed differences between positive (S+) and negative (S−) trials. The latter (random) group, according to this view, would learn to persist in approaching the goal box in the face of frustration in the initial gray-alley task, and such persistence would be the dominant response tendency in the subsequent discrimination whenever r_F was elicited on the S− trials, until the discrimination was learned to the point where avoidance or competing responses predominated over approach tendencies. In contrast, the patterning group, if trained on regular 50% reinforcement in the gray alley long enough to display consistent inhibition of approach responding on N trials, would tend to make (s_F-elicited) competing or goal-box-avoidance responses as soon as r_F was built up on S− trials in the black–white discrimination.

The expected group difference on the S− trials parallels the empirical finding (e.g., Capaldi, 1958; Tyler et al., 1953) that ordinary extinction following extended training under single-alternation 50% reinforcement is more rapid than extinction following a random 50% schedule. Capaldi (1966, 1967) can derive this extinction result from the fact that the dissimilarity between the N lengths of 50% reinforcement and extinction is greater; and hence response decrement in extinction would be greater for a regular-trained group (initially conditioned to a single N length of one) than for a random-trained group (which typically receives many conditionings of N lengths of one, two, or three).

It would appear at first glance that Capaldi's (1967) formulation could just as readily handle the analogous difference with respect to S− trials in discrimination, and thus make a prediction similar to Amsel's (1967). In applying Capaldi's explanation of the extinction result to a typical successive discrimination situation, however, there are obvious complications stemming from the fact that S+ and S− are randomly interspersed in the latter. Capaldi's heavy emphasis upon the cue value of covert stimulus aftereffects would lead us to expect that on the early trials of a black–white discrimination, performance on a given trial by rats which have acquired patterning in a similar setting would be strongly influenced by the stimulus carry-over (S^R or S^N) of the preceding trial, regardless of the external cues (S+ or S−) present on the given or the preceding trial. Whether this influence would be beneficial to the acquisition of the discrimination, relative to acquisition rates of random-trained animals, is not clear. The effect might depend on the particular sequences of S+ and S− trials administered. These could be "biased" so as to produce (spuriously) good performance, as might be expected with sequences approaching single alternation of S+ and S− trials, or poor performance, which might result from sequences containing consecutive S+ or S− trials (e.g., S+S+S+S−S−S−), but it is also possible that as the trial sequences deviated more and more from single alternation of S+ and S− trials they would provide more new information to pattern-trained rats. Further complications ensue from Capaldi's (1967) assumptions that the aftereffects of two or more consecutive N (or R) trials are somewhat dissimilar to single-trial aftereffects, and that the aftereffect of a given trial may include "memories" of stimulus events over and above reward or nonreward experiences. The latter of these assumptions leads to an alternative theoretical strategy in applying Capaldi's formulation to successive discrimination which we discuss later.

With various options available to the sequential theorist, it is apparent that investigating the transfer by pattern- and random-trained groups to successive discrimination would not constitute a crucial test of the Amsel (1967) versus Capaldi (1967) theories. This experiment does, however, provide an opportunity to examine whether Amsel's transfer-mediating mechanism ($r_F - s_F$) applies to an unusual type of transfer situation, and it promises to yield useful information to the sequential theorist regarding the more fruitful strategy in extending Capaldi's instrumental learning theory to discrimination learning.

In our investigation of this transfer problem, we administered discrimination sessions in the form of black- and white-runway trials, the sequences of which were unbiased in the sense that pattern-trained rats would not be expected to show spuriously good or poor performance on them at the outset. The development of patterning in the prior gray-alley task was examined under conditions which controlled for possible sources of nonconditioned patterning, such as the presence of food odors in the goal region on R trials alone (cf. Amsel, Hug, & Surridge, 1969) and differential odor trails left by rats after R and N trials (cf. Morrison & Ludvigson, 1970). Besides providing the main comparison of interest in the discrimination phase, the group receiving random 50% reinforcement in the gray runway served to provide appropriate running orders for the control of odor trails in the prior phase and a means of evaluating the success of this and other precautionary measures.

METHOD

Subjects.—The *S*s were 16 naive male Holtzman rats approximately 120 days old at the start of the experiment.

Apparatus.—Three straight runways were used. Each runway was 10.2 cm. wide and high, and included a 36.7-cm. start box, a 172.1-cm. alley, and a 34.8-cm. goal box. The runway used in the first phase of the experiment was painted gray throughout. The start and goal boxes were delimited by clear Plexiglas doors, and the top of the entire runway was covered by clear Plexiglas. The start boxes of both runways used in the brightness-discrimination phase of the experiment were painted gray. The remainder of one of the runways was painted flat black, and the other runway was painted white. The start-box doors in both alleys were painted gray on the start-box side and the appropriate alley color on the runway side. The goal-box doors and the runway tops were clear Plexiglas. Hunter photorelays and Standard Electric timers were used to measure start, run, and goal times. Start time was measured from the opening of the start-box door to Photobeam 1 located 15.2 cm. beyond the door. Run time was measured from Photobeam 1 to Photobeam 2 located 125.7 cm. further down the alley, and goal time was from Photobeam 2 to Photobeam 3, 58.4 cm. away and 8.9 cm. from the end of the alley. A Plexiglas food cup, 2.54 cm. in diameter and 1.3 cm. deep, was placed behind a 2-cm.-high baffle at the end of the goal box.

Procedure.—The *S*s were reduced to 82% of their ad-lib weight and maintained at this level by a controlled amount of feeding each day throughout the experiment. In the first of 2 pretraining days, *S*s were allowed 5-min. exploration of the alley. On the second day, *S*s were placed in the goal box with the retrace door closed and were allowed to eat 20 45-mg. Noyes pellets from the food cup. The next 36 days constituted the patterning phase of the experiment. During this period, each *S* received three rewarded (R) and three nonrewarded (N) trials per day. The reward consisted of 10 45-mg. Noyes pellets. Half of the *S*s, the regular (REG) group, received the same alternating sequence, RNRNRN, every day. For the remaining *S*s, the random (RAND) group, the following eight sequences were randomly presented across days: RNNRRN, RRNRNN, RNRRNN, RRNNRN, NRRNNR, NNRNRR, NRNNRR, and NNRRNR. On N trials, *S*s were confined to the goal region for 45 sec. On R trials, *S*s were removed immediately after the reward was consumed. The intertrial interval (ITI) of approximately 16 min. was determined by the time necessary for all of the *S*s to complete a trial. The *S*s from each group were alternated in the daily running sequence, so that the reinforcement conditions for any *S* in the running order were inconsistently related to the reinforcement conditions prevailing for the preceding *S* on any given trial. In order to minimize extraneous cues associated with R and N trials, the area on the

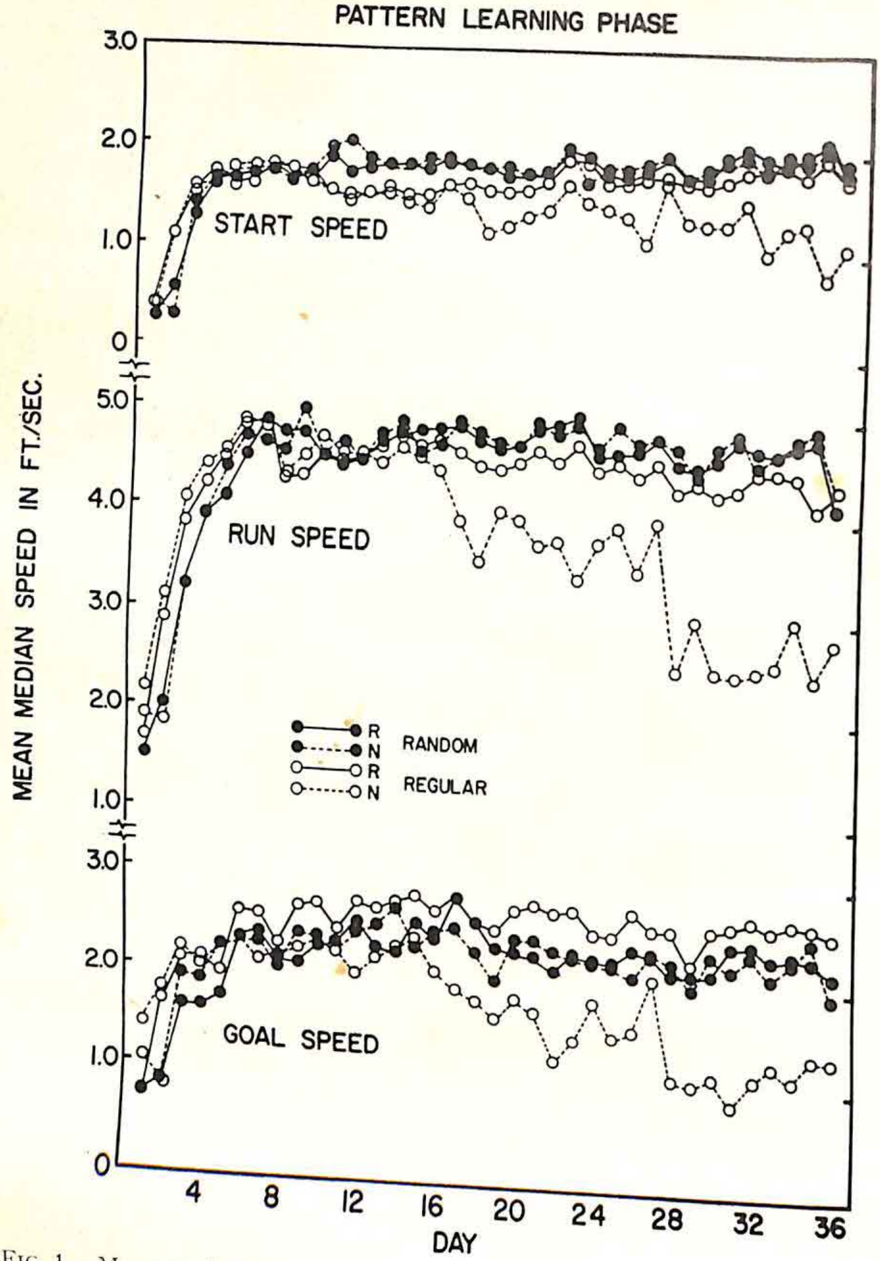


FIG. 1. Mean median speeds on rewarded (R) and nonrewarded (N) trials.

outside of the runway around the goal box was sprinkled with Noyes pellets and powder, and *E* went through the same motions on both types of trials.

On the day following completion of the pattern-learning phase, *S*s were shifted to the black and white runways and given brightness-discrimination training. The eight trial sequences used for the RAND group in the patterning phase were used to program reinforcements for all groups in this phase. Within each of these six-trial sequences, it should be noted, there were two consecutive S+ trials and two consecutive S− trials. The trial sequence used on the first day of discrimination training was RRNNRN for all 16 *S*s. Reward-cue combinations were counterbalanced and were orthogonal to previous reinforcement history. Thus, black was S+ for half of the *S*s in the RAND group and for half of the *S*s in the REG group. Daily running order was the same as in the previous phase. Discrimination training was continued for 14 days. Reward magnitude, goal-box confinement, and ITI were the same as in the pattern-learning phase.

RESULTS

Pattern learning.—Mean median start, run, and goal speeds on R and N trials are presented in Fig. 1. It is clear that Group REG differentiated between R and N trials, whereas group RAND did not. The slower running on N trials by the regular group developed gradually and was eventually evident in all three sections of the runway. Analysis of variance on the data from Days 31–36 provided statistical verification of these effects. The N-trial speeds of Group REG were significantly lower than R-trial speeds and significantly lower than both R- and N-trial speeds of the RAND group: Schedule × Reward interaction—start speed, F (1, 14) = 11.34; run speed, F (1, 14) = 60.32; goal speed, F (1, 14) = 184.58; all $ps < .01$.

The patterning behavior apparent in Fig. 1 did not develop uniformly within the daily sessions. Differential responding appeared initially in the first two trials of the daily session and was relatively retarded in the middle two trials of each session. These effects are clear in Fig. 2, which shows the run speeds of the REG group at five stages of acquisition. In this figure, the data are presented in terms of the first R and first N trial, etc. Statistical analyses of these data indicated the following. The N-trial speeds were uniformly faster than R-trial speeds on Days 1 and 2, F (1, 7) = 9.25, $p < .05$; there were no significant differences on Days 8 and 9; on Days 17 and 18 the first N trial was slower than the first R trial, F (2, 14) = 4.67, $p < .05$, but other pairs of R and N trials did not differ significantly by Fisher's *lsd* test with $p = .05$; on Days 26 and 27 both the first and third N trials were significantly slower than the first and third R trials and, in addition, the first N trial was significantly slower than all other trials, F (2, 14) = 11.86, $p < .01$, followed by Fisher's *lsd* test; and finally, on Days 35 and 36 N-trial speeds were uniformly slower than R-trial speeds, F (1, 7) = 35.26, $p < .01$. The acquisition data obtained with the start- and goal-speed measures were similar in form to the run-speed data.

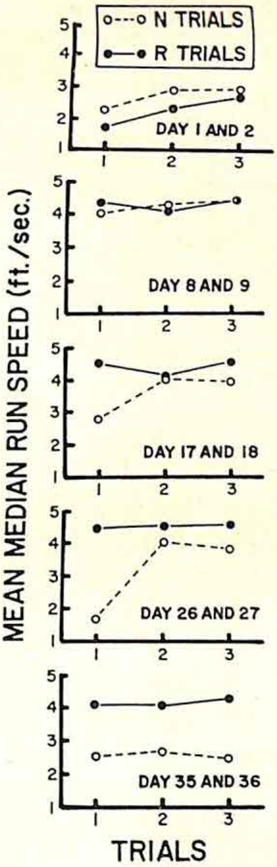


FIG. 2. Mean median run speeds on R and N trials at five stages of acquisition. (Trial 1 refers to the first R and first N trial, etc. The trial sequence was always RNRNRN.)

Brightness discrimination.—The performance of the two groups in the brightness-discrimination phase is presented in Fig. 3. It is clear that the group with a history of regularly alternating 50% reinforcement demonstrated considerably more differential responding than the group with the random 50% reinforcement history. It is also evident that this greater discrimination on the part of Group REG was due primarily to slower running in the $S-$ alley. The differences were significant in all three sections of the runway: Schedule × Reward interaction—start speed, F (1, 12) = 30.50; run speed, F (1, 12) = 13.02; goal speed, F (1, 12) = 11.79; all $ps < .01$. The difference in $S+$ speeds in the goal section did not attain statistical significance by Fisher's *lsd* test ($p = .05$).

Figure 4 presents the trial-by-trial mean run speeds of the two groups on the last session of the patterning phase and the first five sessions of the black–white dis-

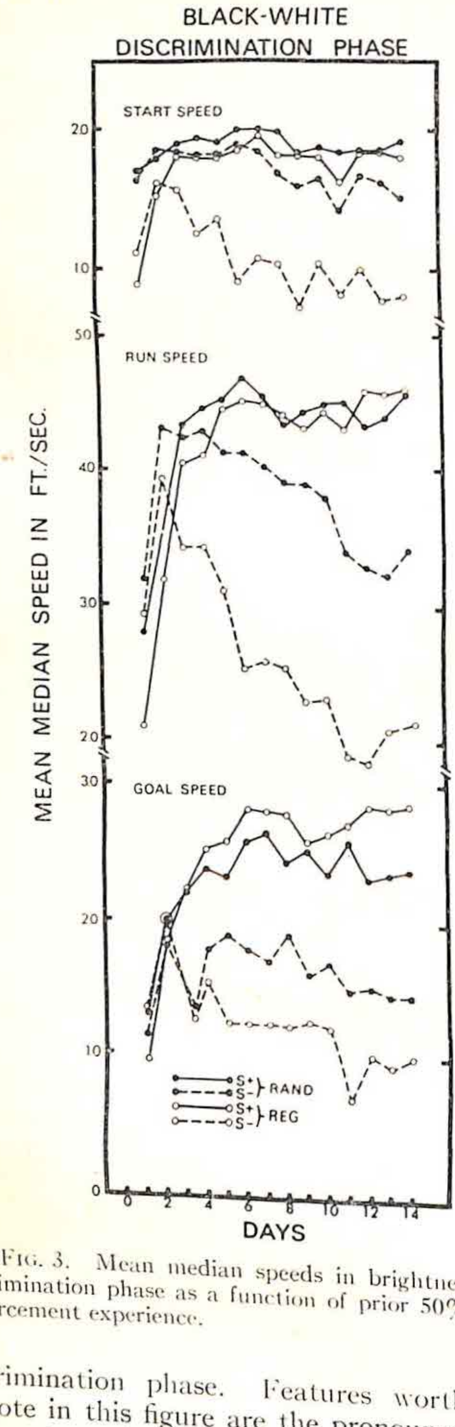


FIG. 3. Mean median speeds in brightness discrimination phase as a function of prior 50% reinforcement experience.

crimination phase. Features worthy of note in this figure are the pronounced depressions of run speeds in both groups on the initial S+ and S− trials in the first discrimination session (D-1), the absence of trial-to-trial changes which reflect prior pattern learning in the REG group, the striking similarity of the two groups with respect to trial-to-trial transitions in run speed generally, and the suggestion (D-3) of faster development of the discrimination by the REG group. Individual plots were smooth enough to determine the crossover points after which S+ run speeds consistently exceeded S− run speeds. These crossover points occurred prior to D-4 in six of the REG rats and three of the RAND rats, a difference which is not statistically significant.

DISCUSSION

Patterning.—It is unlikely that the stimulus control apparent in Group REG was dependent on ancillary stimuli such as odors from food in the goal region or residual cues left by *S*s prior in the running sequence. Stimuli from food in the goal region on R trials should have affected both groups equally, yet Group RAND showed no evidence of patterning, even in the goal region. The alternation of RAND and REG *S*s in the daily running sequence meant that there was no consistent relationship in the trial-reinforcement events prevailing for any two consecutive *S*s, and thus there was no possibility that any residual stimuli left by one *S* could be used to predict reward by the next *S* in the sequence. It seems likely that the source of stimulus control for the REG group was the memory or aftereffects of the reward conditions prevailing on the previous trial. In Capaldi's (1967) terms, S^R becomes associated with inhibition because R trials are always followed by N trials, and S^N acquires instrumental habit strength because N trials are always followed by R trials. Thus, *S* should run slower when S^R is the prevailing stimulus, that is, on N trials.

While Capaldi's model provides a vehicle for describing patterning behavior in conditioning terms, there are some aspects of the present data that are not completely compatible with sequential theorizing. Capaldi (1967) assumes that the stimulus aftereffects of reward and nonreward are highly responsive, that is, "the stimulus trace occurring on a given trial is totally determined by the reward outcome on the previous trial [pp. 71–73]." In the context of the present experiment, this means that the stimuli controlling patterning behavior in the REG group should be either simple S^Ns or simple S^Rs or, in other words, the S^Rs controlling behavior on Trials 2, 4, and 6 should be all the same because each of these N trials is preceded by a single R trial. This implies that stimulus control

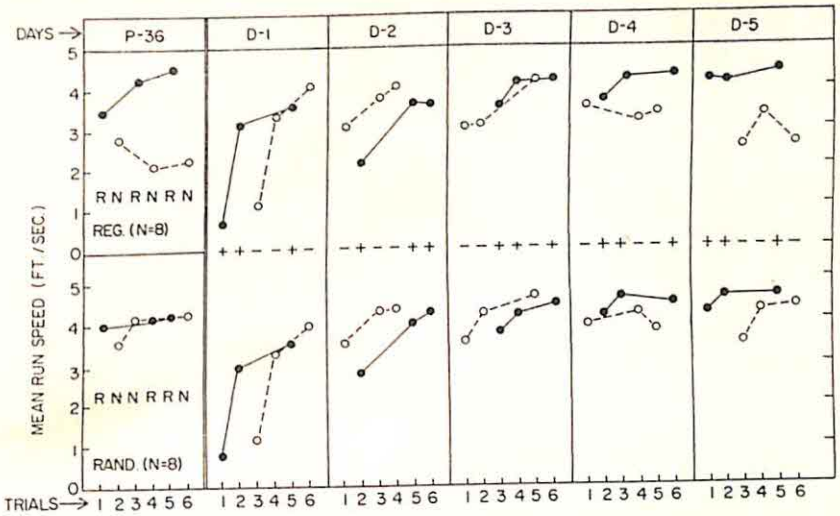


FIG. 4. Mean run speeds of the REG (upper panels) and RAND (lower panels) groups on individual rewarded (filled circles) and nonrewarded (open circles) trials in the last session of the patterning phase (P-36) and the first five sessions of the discrimination phase (D-1 through D-5).

should develop at an equal rate at all positions in the trial sequence. However, the present data clearly show that the development of patterning behavior was quite different on Trial 2 from that on Trials 4 and 6.

Our interpretation of these data is that the stimulus controlling behavior on Trial 2 was more distinctive than the stimuli controlling behavior on Trials 4 and 6. The distinctiveness scale developed by Murdock (1960) seems to be relevant to these data. On Murdock's scale, the distinctiveness of any single stimulus from a series of stimuli is found by obtaining the sum of the log energy values by which that stimulus is separated from each other stimulus in the series. A percent distinctiveness value (*D*%) can then be obtained for each stimulus. This is done simply by taking the distinctiveness value found for any one stimulus and forming a ratio with the sum of the distinctiveness values found for all stimuli in the series. Murdock found that this scale accurately predicted errors in judgment among stimuli that differed in intensity levels and, more importantly for the present experiment, the scale was also found to be applicable to serial learning tasks, both human and animal. In order to apply the scale to serial learning phenomena, it is necessary to make two assumptions. First one must assume that the distinctiveness of a given stimulus item depends upon its position in the sequence and, second, one must assume that the ordinal continuum of the stimuli is logarithmic. Given these assumptions, Murdock found that *D*% accurately described the bowed serial position curve found in human rote learning as well as the error distribution in some rat maze learning. The application of the *D*% scale to the present data requires the additional assumption that there is a stimulus value associated with the ordinal position of each trial in a sequence. With these assumptions, the *D* and *D*% values relevant for a 6-trial sequence were calculated as indicated by Murdock (1960, p. 17) and are presented in Table 1. The *D* values in the table were obtained in the following way. The log values corresponding to the position of each trial in the 6-trial sequence were used to construct a 6 × 6 matrix. Entries above the principle diagonal were obtained by multiplying the log unit value of each column by 1, and entries below the principle diagonal were obtained by multiply-

TABLE 1

DISTINCTIVENESS (*D*) AND PERCENT DISTINCTIVENESS (*D*%) VALUES FOR EACH TRIAL OF A SIX-TRIAL SEQUENCE

Trial	Log trial position	*D*	*D*%
1	.0000	2.8574	27.2
2	.3010	1.6534	15.7
3	.4771	1.3012	12.4
4	.6021	1.3012	12.4
5	.6990	1.4950	14.1
6	.7782	1.9118	18.2

Note.—$\Sigma D = 10.5200$.

ing the log unit value of each column by -1. The principle diagonal was obtained by multiplying the log values of each column by the term $-[n-(2i-1)]$, where n equals 6 and i equals the number of the particular trial in the sequence 1 through 6. The D values for each row (see Table 1) were obtained by summing the cell entries across columns. The $D\%$ values for each trial were obtained from the ratio $D/\Sigma D \times 100$. The $D\%$ values show that the first trial of the day should be the most distinctive trial, the last trial the second most distinctive, etc.

These distinctiveness values are applicable to the patterning data in the following way. As discussed above, the stimulus controlling patterning behavior is presumed to be the aftereffect of the reward conditions prevailing on the previous trial in the sequence. Therefore, the effectiveness of the $D\%$ values should be delayed one trial. Applying the $D\%$ values to N-trial speeds, we see that the stimulus (S^{R_1}) controlling running on Trial 2 (the first N trial) should be considerably more distinctive than the reward aftereffects controlling running on Trials 4 and 6. We also see from Table 1 that the aftereffect of the third R trial should be slightly more distinctive than the aftereffect of the second R trial. It is apparent from Fig. 2 that the development of differential running speeds was in agreement with what would be expected on the basis of Murdock's (1960) distinctiveness scale. Differential responding developed much earlier on Trial 2 than on Trials 4 and 6. There was also a very slight tendency for the discrimination to develop more rapidly on Trial 6 than on Trial 4.

These data support an interpretation of patterning behavior that is not consonant with Capaldi's (1967) assumption of high responsivity for S^R and S^N. Instead, the aftereffects occurring later in the trial sequence appear to be less distinctive than those occurring earlier in the sequence. A reasonable hypothesis concerning the factors responsible for the differential distinctiveness would seem to be the following. If the stimulus aftereffects of reward and nonreward are teated as memories and, like other memories, are subject to interference (cf. Spear, 1967), then it is reasonable to assume that within-session proactive interference should make the S^Rs and S^Ns later in the trial sequence less discriminable than those earlier in the sequence. For example, an *S* at the start of Trial 2 has had only one previous trial that day, an R trial, and has only to recall S^R to demonstrate "appropriate" behavior, whereas an *S* at the start of Trial 4 has had two previous R trials and one previous N trial and now must recall which type of trial came last. Mackintosh (1970) has recently found that proactive interference between R and N trials serves as a useful explanation for some of the effects of intertrial interval and trial sequence on the partial reinforcement extinction effect.

Discrimination.—The black–white discrimination data were well in line with expectations from Amsel's (1967) theorizing. According to this view, the large difference between the groups in speeds on S− trials would result from the $(r_F - s_F)$-mediated transfer of both persistence in approach in the RAND group and avoidance or competing-response tendencies in the REG group. Transfer of either type from the gray-runway situation would not be expected to exert appreciable influence in the black–white discrimination until *de novo* buildup of r_F occurred on the S− trials, to magnitudes approaching those in the latter part of the gray-runway task. Elicitation of weaker r_Fs on the earliest S− trials of the discrimination would, in fact, be expected to produce a mild energizing of approach in both groups, as was the case on the second day of the discrimination (see Fig. 4, Session D-2; higher S− than S+ run speed was shown by 12 of the 16 rats on D-2). Thus, while the Amsel theory predicts a large difference in terminal S− speeds, it does not seem to predict any substantial difference between REG and RAND groups in *rate* of discrimination, which also accords with the present data.

The trial-by-trial changes in speed at the outset of discrimination (Fig. 4) provided no evidence that response tendencies attached to aftereffects of rewarded and nonrewarded trials ($S^R \rightarrow$ avoidance, $S^N \rightarrow$ approach) in the REG group carried over from the patterning to the discrimination phase. After the initial decrements in speed (Trials 1 and 3, D-1) produced by the shift from gray to black and white alleys, the REG rats showed little tendency to inhibit responding on the second of two consecutive S+ trials (e.g., Trial 2 of D-1, Trial 6 of D-2) or on S− trials following S+ trials (e.g., Trial 6 of D-1, Trial 3 of D-2), and the behavioral transitions of the two groups were nearly identical. This result might be interpreted as indicating that aftereffects included specific memories of the stimulus setting (gray, black, and white cues) as well as traces of reward and nonreward (cf. Capaldi, 1967, p. 116) and that a large degree of generalization decrement operated to remove carryover of $S^N \rightarrow$ approach or $S^R \rightarrow$ avoidance tendencies from gray- to black- or white-alley trials.

With greater emphasis on the inclusion of stimulus events above and beyond goal events in the total aftereffect of a trial, Capaldi's (1967) formulation becomes applicable to successive discrimination in a manner that is compatible with the main finding of this study, the superior black–white discrimination by the REG group. This application can proceed on the assumptions that aftereffects are largely nullified in their next-trial influence (but "placed into storage") whenever a transition in external cues (e.g., black to white) takes place in the trial schedule and that the aftereffect (memory) of a given trial carries over instead to the next trial having the same external cue, the reappearance of this cue having the function of reinstatement or "retrieval from storage." Thus, for example, in the trial sequence B−W+W+B−, the aftereffect of the first trial (S^N) does not carry over into the second trial because of the transition from black (B) to white (W), but does carry over to the fourth trial when the cue of the first trial (B) next reappears, S^N being reinstated at the outset of the fourth trial. With this approach, the S− trials can be treated in isolation from the S+ trials, as if the former were an uninterrupted extinction series, and Capaldi's prediction regarding the effects of prior pattern-learning on black–white would parallel the simple extinction case, in agreement with Amsel's (1967) prediction and the present results.

REFERENCES

AMSEL, A. Partial reinforcement effects on vigor and persistence. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* New York: Academic Press, 1967.

AMSEL, A., HUG, J. J., & SURRIDGE, T. Subject-to-subject trial sequence, odor trials, and patterning at 24-h ITI. *Psychonomic Science,* 1969, **15,** 119–120.

AMSEL, A., & WARD, J. S. Frustration and persistence: Resistance to discrimination following prior experience with the discriminanda. *Psychological Monographs,* 1965, **79** (4, Whole No. 597).

CAPALDI, E. J. The effect of different amounts of training on the resistance to extinction of different patterns of partially reinforced responses. *Journal of Comparative and Physiological Psychology,* 1958, **51,** 367–371.

CAPALDI, E. J. Partial reinforcement: A hypothesis of sequential effects. *Psychological Review,* 1966, **73,** 459–477.

CAPALDI, E. J. A sequential hypothesis of instrumental learning. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* New York: Academic Press, 1967.

CAPALDI, E. J., & STANLEY, L. R. Temporal properties of reinforcement aftereffects. *Journal of Experimental Psychology,* 1963, **65,** 169–175.

MACKINTOSH, N. J. Distribution of trials and the partial reinforcement effect in the rat. *Journal of Comparative and Physiological Psychology,* 1970, **73,** 341–348.

MACKINTOSH, N. J., & HONIG, W. K. Blocking and enhancement of stimulus control in pigeons. *Journal of Comparative and Physiological Psychology,* 1970, **73,** 78–85.

MORRISON, R. R., & LUDVIGSON, H. W. Discrimination by rats of conspecific odors of reward and nonreward. *Science,* 1970, **167,** 904–905.

MURDOCK, B. B., JR. The distinctiveness of stimuli. *Psychological Review,* 1960, **67,** 16–31.

SPEAR, N. E. Retention of reinforcer magnitude. *Psychological Review,* 1967, **74,** 216–234.

THOMAS, D. R., FREEMAN, F., SVINICKI, J. G., BURR, D. E. S., & LYONS, J. The effects of extradimensional training on stimulus generalization. *Journal of Experimental Psychology,* 1970, **83** (1, Pt. 2).

TYLER, D. W., WORTZ, E. C., & BITTERMAN, M. E. The effect of random and alternating partial reinforcement on resistance to extinction in the rat. *American Journal of Psychology,* 1953, **66,** 37–65.

(Received February 23, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 10–16

SERIAL POSITION EFFECTS FOR REPEATED FREE RECALL: NEGATIVE RECENCY OR POSITIVE PRIMACY?[1]

WAYNE H. BARTZ,[2] MARION Q. LEWIS, AND GENE SWINTON

Iowa State University

Repeated free recall was investigated in two experiments. In each experiment, lists of words were recalled immediately after presentation and again after all lists had been presented (a final free recall, FFR). Lists were presented at two rates in Exp. I and FFR followed each of two blocks of 10 lists. In Exp. II, one group of *S*s recalled each list twice at initial presentation and a second group recalled each list once at initial recall. A third group recalled each list once but was informed in advance about the FFR. Results showed that opportunity for rehearsal (presentation rate) increased performance at initial recall and the FFR. Anticipation of a delayed (i.e., repeated) test of memory also increased performance at the FFR, principally for words presented in early serial positions. FFR serial position effects showed lowest recall at terminal serial positions. This finding was interpreted as a positive effect of processing items in long-term memory (positive primacy) rather than a short-term memory effect (negative recency).

Craik (1970) reported a "negative recency" effect for repeated free recall. His *S*s were presented a series of lists for free recall. After recall of the last list, *S*s were required to recall words from all lists presented, a final free recall (FFR). Serial position effects for the FFR showed a lower level of recall for recent items than for items presented in earlier serial positions. This negative recency finding was interpreted as a short-term memory (STM) effect. That is, recent list items were coded in a temporary STM, and earlier items were processed into a more permanent, or long-term memory (LTM). Recent list items tended to be retrieved first in the initial recall, and Craik concluded that these items were less effectively encoded in LTM. However, *S*s in the Craik study were tested in four separate experimental sessions, each requiring the recall of 10 lists and an FFR. While *S*s were unlikely to have anticipated the FFR in the first session, it is likely that they anticipated it in the remaining sessions. The present study investigates the possibility that anticipation of the FFR influences information-processing strategies.

Cohen (1970) has also used the FFR method but did not find marked negative recency effects; serial position effects in FFR showed decreasing recall across input positions. Cohen suggested that this finding was due to the differential rehearsal of items; items in early input positions tend to be rehearsed to a greater extent than items in late input positions (Rundus & Atkinson, 1970). Even though there were repeated FFR tests in some of Cohen's experiments, the effects of repeated testing were not assessed.

Experiment I presented two blocks of 10 lists, each followed by an FFR. If anticipation of the FFR test influences *S*'s information-processing strategy, performance should differ for the two FFR tests. In addition, two presentation rates were used in order to provide different amounts of time for rehearsal in STM (Cohen, 1970). It was expected that FFR performance would be higher with slower presentation rates.

EXPERIMENT I

Method

Subject.—The *S*s were 15 volunteers enrolled in psychology courses at Iowa State University. Eight

---

[1] This research was supported by United States Public Health Service Research Grant MH-13192 from the National Institute of Mental Health.

[2] Requests for reprints should be sent to Wayne H. Bartz, Department of Psychology, Iowa State University, Ames, Iowa 50010.

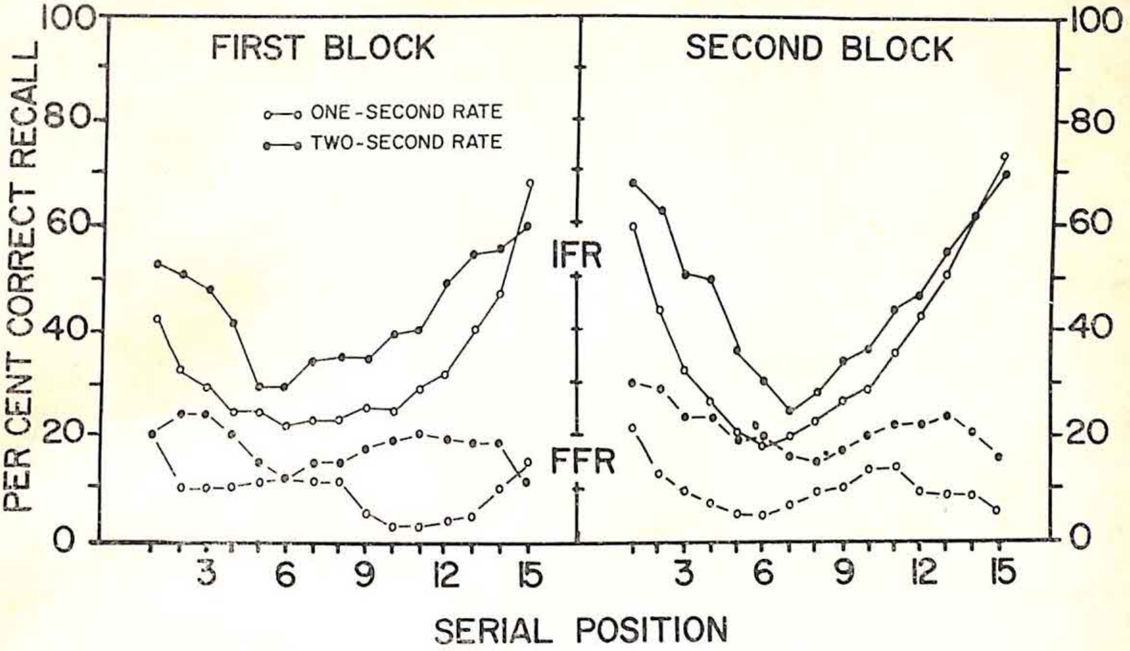


FIG. 1. Serial position effects for Exp. 1.

of the *S*s received extra course credit for participation and 7 were paid.

Lists.—Twenty 15-word lists of Thorndike and Lorge A and AA words were presented on a Lafayette 23017 memory drum. The lists were presented in two blocks of 10 lists each. Within each block, 5 lists were presented at a 1-sec. rate, 5 at a 2-sec. rate. The order of the rates was random within each block. The recall interval was 30 sec. An FFR followed each block of lists. There was no time limit on the FFR.

Procedure.—The *S*s were tested individually or in pairs. The lists were re-ordered for each *S* or pair of *S*s. They recalled in writing, covering each word with a card as they wrote. A new page in the recall booklet was used for each list. The *S*s were not informed in advance about the FFR that followed each block of 10 lists.

Analysis.—Serial position effects were computed for initial recall (IFR) and FFR in each block of lists. For analysis, lists were divided into segments based on serial position (SP): first (SPs 1–5), middle (SPs 6–10), and last (SPs 11–15). The number of words recalled at first recall and the FFR were analyzed by a 2 × 2 × 2 × 3 × 15 (Blocks × Rates × Recalls × Segments × *S*s) analysis of variance.

Results

Because none of the interactions of *S*s with other factors were significant ($p < .05$), these interactions were pooled and used as the denominator in the F tests. As in the Craik (1970) and Cohen (1970) studies of repeated free recall, performance was markedly higher at the IFR (41%) than at the FFR (15%), F (1, 322) = 469.43, $p < .01$. Recall differed significantly for the list segments, F (2, 322) = 42.31, $p < .01$; recall was lower for the middle segment (21%) than for the first and last (31% and 32%) segments, which did not differ. Performance for list segments was different at the IFR and the FFR, F (2, 322) = 29.29, $p < .01$. Performance on the three segments at the FFR was about the same (19%, 14%, and 14% for first, middle, and last segments), but performance at the IFR showed marked primacy and recency (42%, 28%, and 51%).

Recall was higher in the second block (30%) than in the first block (26%), F (1, 322) = 12.48, $p < .01$. Recall was higher at the 2-sec. rate (33%) than at the 1-sec. rate (23%), F (1, 322) = 81.38, $p < .01$. The interaction of rate and list segments was significant, F (2, 322) = 2.73, $p < .05$. The percentages of correct recall for the three segments at the 2-sec. rate were 37, 25, and 37 and 24, 17, and 28 at the 1-sec. rate. No other tests in the analysis of variance were significant.

Serial position effects are shown in Fig. 1. The curves have been smoothed by averaging performance at each SP, except for the first and last, with performance at adjacent SPs. The significant effects de-

TABLE 1

Proportion of Total First-Block Intrusions in the Second-Block Final Free Recall

Present Rate	List segment			
	First	Middle	Last	Total
1 sec.	.125	.137	.050	.312
2 sec.	.358	.102	.228	.688
Total	.483	.239	.278	1.000

scribed above are obvious in Fig. 1. The FFR serial position curves showed negative recency only in the second block of lists.

The number of intrusions in IFR was approximately the same at the two blocks of lists and for the two rates. The intrusion rate was slightly less than one-half of a word per *S* per list. There was a markedly higher intrusion rate in the second-block FFR than in the first-block FFR. The rate of intrusions per *S* was 3.60 for the first block and 11.07 for the second. First-block intrusions were all extraexperimental, but in the second-block FFR 51% were words presented in the first block. Table 1 shows the distribution of the experimental intrusions; two-thirds came from lists presented at the slow rate and nearly half of them were words presented in the first list segment (SPs 1–5).

Discussion

The results of the present experiment will also be discussed within the context of Exp. II. The greater recall at the slower presentation rate may be interpreted as a function of differential rehearsal. The slower rate enabled more rehearsal, increasing item retrievability (Cohen, 1970). The high rate of intrusions in the second-block FFR may have attenuated that performance, particularly for early serial positions because most intrusions were from early SPs in first-block lists. There appeared to be some "learning to learn" in the experiment; performance was higher for the second block of lists. The general shape of the serial position curves also changed over blocks, showing greater primacy and recency in initial recall (IFR) and a higher level of performance (with negative recency) in the second block. Overall, the results suggest that learning to learn is principally a matter of developing more efficient LTM encoding and retrieval strategies. On the other hand, the improvement over blocks may have been due to anticipation of the FFR. That is, *S*s could have expected the FFR in the second block, and this expectation may have influenced performance. This possibility was further investigated in Exp. II

Experiment II

The results of Exp. I, including the intrusion findings, suggest that anticipation of the FFR increases the recall of items from early serial positions. To investigate this possibility, two groups of *S*s recalled 10 lists and were then given an FFR test of memory. One of the groups was informed in advance that there would be an FFR test. This method avoids the problem of intrusions in FFR.

Lewis (1971) has also studied repeated free recall and interpreted the results, as did Craik (1970), in terms of coding in STM and LTM. Her *S*s performed two immediately successive recalls of lists. Results showed first and second recalls nearly identical at early serial positions, but there was a sharp decline between recalls for recent items. There was no negative recency in the Lewis study. Craik's FFR, however, followed the presentation and recall of several lists; in the Lewis study, the second recall immediately followed the first (there was no FFR).

Some investigators have found that test trials have the same effects as presentation (e.g., Lachman & Mistler, 1970). In the present experiment, a third group, as in the Lewis (1971) study, was required two immediately successive recalls of each list, and then an unanticipated FFR. If recall serves as a second presentation, then this group should show a higher level of recall at the FFR. It might also be predicted that this repeated recall manipulation would differentially affect items in early serial positions. This would follow from the assumption that items in early serial positions would be retrieved twice from LTM while late items would be recalled from STM at the first recall, from LTM at the second recall.

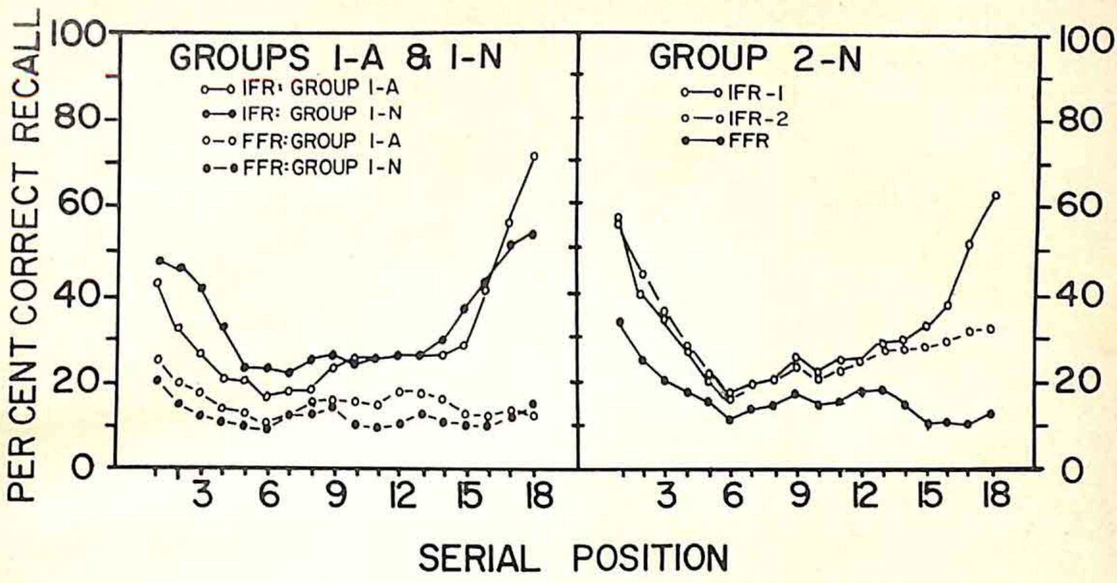


FIG. 2. Serial position effects for Exp. II.

Method

Subjects.—The *S*s were 45 volunteers enrolled in psychology courses at Iowa State University. They received extra course credit for participation.

Procedure.—Ten 18-word lists of Thorndike and Lorge A and AA words were presented at a 2-sec. rate on a Lafayette 23017 memory drum.

The *S*s were assigned to treatment groups in order of their appearance at the laboratory. The *S*s in Groups 1-N and 2-N (no anticipation) were not informed in advance about the FFR. Group 1-A (anticipation) was told that there would be a "final exam" on all words presented. Group 1-N and 1-A were required to recall each list once; Group 2-N was required to recall each list twice, the second recall immediately following the first. The recall intervals were 30 sec. In all other respects, the procedure was identical to that for Exp. I.

Analysis.—Serial position effects were computed for the recalls in each group. For analysis, lists were divided into segments based on serial position: first (SPs 1–6), middle (SPs 7–12), and last (SPs 13–18). The number of words recalled at first initial free recall (IFR-1) and the FFR were analyzed by means of a 3 × 3 × 2 (Groups × Segments × Recalls) analysis of variance. Conditional probabilities for the FFR were also computed.

Results

Serial position effects are shown in Fig. 2; the curves for groups with one and two initial recalls are separated simply for purposes of clarity. The curves have been smoothed by averaging performance at each SP, except for the first and last, with performance at adjacent SPs. The curves for initial recall in Group 2-N replicate precisely the findings of Lewis (1971); IFR-2 is lower than IFR-1 only at recent serial positions. Negative recency is apparent in Groups 2-N and 1-A; the FFR curve for Group 1-N is relatively flat across all serial positions. In comparison with the curves for Exp. I (Fig. 1), Groups 1-A and 2-N show serial position curves similar to those for the second block; the Group 1-N curve resembles the curve for the first block of lists in Exp. I.

IFR-1 was greater than FFR, F (1, 42) = 500.34, $p < .01$; percentages were 32% for IFR-1 and 15% for FFR. Segments differed, reflecting serial position effects shown in Fig. 1, F (2, 84) = 19.96, $p < .01$. Recall for first and last segments did not differ (25% and 27%), but recall for these segments was greater than that for the middle segment (19%). Recalls and segments interacted, F (2, 84) = 55.70, $p < .01$, such that the last segment was higher than the middle at IFR-1 (recency peak), but the segments did not differ at FFR.

Groups did not differ significantly in terms of overall recall ($F < 1$), but there was a Groups × Recalls interaction, F (2, 42)

TABLE 2

CONDITIONAL PROBABILITY OF FINAL FREE RECALL BASED ON INITIAL RECALL

Group	Probability	Item type	
		LTM[a]	STM[b]
1-N	$P(C_{IFR\text{-}1})$	.30	.41
	$P(C_{FFR})$	.13	.12
	$P(C_{FFR}/C_{IFR\text{-}1})$	.35	.24
1-A	$P(C_{IFR\text{-}1})$	.24	.40
	$P(C_{FFR})$	.17	.14
	$P(C_{FFR}/C_{IFR\text{-}1})$	.51	.29
2-N	$P(C_{IFR\text{-}1})$	.28	.42
	$P(C_{FFR})$	.18	.14
	$P(C_{FFR}/C_{IFR\text{-}1})$	.50	.27
	$P(C_{IFR\text{-}2})$	.28	.31
	$P(C_{FFR}/C_{IFR\text{-}2})$	.53	.36
	$P(C_{FFR}/C_{IFR\text{-}1+IFR\text{-}2})$	.54	.35

[a] Serial Positions 1–12.
[b] Serial Positions 13–18.

= 5.88, $p < .05$. At the initial recall, Group 1-A recalled fewer words (30%) than did Groups 1-N (34%) and 2-N (33%). Specifically, Fig. 2 shows that Group 1-A recalled fewer words from early SPs. At the final recall, Group 1-N recalled fewer words (12%) than did Groups 1-A (15%) and 2-N (17%). No other effects or interactions were significant.

Because recall differed for the groups at initial recall, forgetting scores (difference between IFR-1 and FFR) were analyzed by a 3 × 3 (Groups × Segments) analysis of variance. Groups differed significantly F (2, 42) = 4.74, $p < .05$. Tests among group means by the Newman-Keuls procedure showed that forgetting was greater in Group 1-N (21%) than in Groups 1-A (15%) and 2-N (16%). Segments also differed, F (2, 84) = 46.86, $p < .01$; there was more forgetting for the last segment (28%) than for first (14%) and middle (9%) segments. The Groups × Segments interaction was marginally significant, F 4. 84) = 1.39, $p < .10$. Groups did not differ in terms of forgetting at the last segment, and Group 1-N showed a higher level of forgetting than did the other groups at first and middle segments.

Probabilities and conditional probabilities of recall are given in Table 2. For analysis, words in SPs 1–12 (first and middle list segments in the preceding analyses) were termed LTM items, and words in SPs 13–18 (last list segment) were termed STM items. As in Craik's (1970) analysis of conditional probabilities, STM probabilities were lower than LTM probabilities in all groups. The present conditional probabilities for LTM items show that the probability of correct recall at the FFR, given correct recall at IFR-1, was higher in Groups 1-A and 2-N than in Group 1-N. The STM conditional probabilities based on IFR-1 recall were approximately the same for all groups. However, the probability of correct recall, given correct recall at IFR-2 (in Group 2-N), was almost identical to the LTM conditional probability for Group 1-N.

Discussion

The lack of negative recency Group 1-N does not replicate the Craik (1970) findings. It is likely that *S*s in Group 1-N did not anticipate that items would have to be recalled more than once. However, *S*s in Group 2-N were aware that there would be a delayed (repeated) test of memory. These *S*s, then, may have encoded items somewhat differently, enabling retrieval at a later time. This possibility may also bear on the Craik findings. That is, the negative recency reported by Craik may not have been so much a negative effect for recent items as a *positive* effect for *early* items. That is, in Group 2-N in the present experiment, the repeated recall of lists facilitated recall of early items at the FFR. When *S*s were explicitly informed that there would be an FFR (Group 1-A), final recall showed higher retention for items in early SPs, a positive primacy effect.

The high FFR performance on Group 2-N could be interpreted as an increase in item strength due to initial recall. This apparent increase in item strength as a function of two prior recalls is compatible with Lachman and Mistler's (1970) observation that a recall or test trial has the same effect as a presentation. This increase in term strength, however, affected only LTM items; Groups 2-N and 1-N differed only at early SPs. At late SPs, these groups did not differ in FFR performance. However, the recall of late items in Group 2-N might be construed as *one* recall from LTM. That is, the first recall might represent STM and the second recall a retrieval from LTM. If this were the case, late-item performance in Group 2-N should be equivalent to the LTM performance of Group 1-N, which had one retrieval from LTM. The conditional probabilities tend to support this interpretation. The LTM (early-item) conditional probability on FFR was .50 for Group 2-N, which recalled each list twice; the LTM conditional probability was only .35 for Group 1-N, which recalled each list once before the FFR test. A different pattern was observed for STM (recent items). The FFR conditional probabilities for the two immediate recalls of Group 2-N were .27 for IFR-1 and .35 for IFR-2. That is, the conditional probability for IFR-1 was similar to that for other groups (Group 1-N, .24 and Group 1-A, .29). The higher conditional probability for STM (recent items) at IFR-2 was the same as the LTM conditional probability of Group 1-N.

The present findings offer only equivocal support for this interpretation. If the second recall in Group 2-N represents retrieval from LTM, there should be a higher level of performance at the FFR for these items due to their increased item strength. However, late-item (STM) FFR performance and forgetting level for this group did not differ from the other groups. Alternatively, the finding of no differences among groups for recent items may be considered an STM effect (Atkinson & Shiffrin, 1968; Cohen, 1970). Items presented in late SPs cannot be encoded in LTM because the rehearsal and encoding is interrupted by recall. The interpretation of increased item strength due to initial recall is also not supported in that Group 1-A, which recalled each list but once, showed an LTM conditional probability on FFR, which was the same as that for Group 2-N.

Results of Exp. II suggest that the *anticipation* of a later recall affects the nature of encoding an item in memory, enabling retrieval at a later time. Group 1-A was explicitly informed at the beginning of the experiment that there would be an FFR, and the performance of these *S*s differed from that of the *S*s in the other groups. Primacy was lower in initial recall (Fig. 2), perhaps because *S*s attempted to encode additional long-term retrieval cues for items, requiring more processing attention per item and reducing the number of items that could be processed. This group of *S*s showed the least forgetting between initial and final recall. The general shape of the serial position curve for Group 1-A was quite similar to the second-block curve in Exp. I. In contrast, the curve for Group 1-N was similar in shape (lower primacy and recency effects) to the curve for the first block of lists in Exp. I.

It was suggested earlier that the difference in the shape of the curves of Exp. I was due to more efficient strategies of LTM encoding and retrieval. The results of Exp. II indicate that the improvement over blocks and the shape of the serial position curves in Exp. I reflect effects of anticipation of the FFR. In other words, the results of Groups 1-A and 1-N, in which the FFR was and was not expected, replicate the findings for blocks in Exp. I.

The LTM conditional probability for Groups 1-A and 2-N (Table 2) were approximately equal and were higher than the LTM probability for Group 1-N. The higher LTM conditional probability may be attributed to *S*'s anticipation of a later recall test. Group 1-A was informed about the FFR and Group 2-N was instructed that each list would have to be recalled twice. Although Group 2-N was not informed about the FFR, these *S*s did know that a delayed (30-sec.) test of memory was required. The FFR serial position effects showed positive primacy in Groups 1-A and 2-N, the groups instructed that a later recall of items would be required.

The effect of anticipation on FFR performance was specific to items presented in first and middle serial positions. According to several models of memory (e.g., Atkinson & Shiffrin, 1968), such items could be processed into an LTM. In contrast, items at the end of the list, which were not affected by advance knowledge in the present experiments, could be processed only into an STM characterized by limited capacity and rapid decay. The results of the present study are consistent with this interpretation, which has been employed by other investigators of repeated free recall (Cohen, 1970; Craik, 1970; Lewis, 1971).

Present results indicate that LTM items are influenced by the opportunity for rehearsal (i.e., presentation rate) and anticipation of a later memory test. Taken together, the results of the present experiments indicate that knowing in advance about a delayed test of memory, even a test delayed just 30 sec., brings about changes in memory processing that are not observed when this knowledge is absent. Thus the negative recency reported by Craik may have been in actuality more of a "positive primacy" effect because anticipation of the FFR was not controlled.

REFERENCES

ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 2. New York: Academic Press, 1968.

COHEN, R. L. Recency effects in long-term recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 672–678.

CRAIK, F. I. M. The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 143–148.

LACHMAN, R., & MISTLER, J. L. Rehearsal, test trials, and component processes in free recall. *Journal of Experimental Psychology*, 1970, **85**, 374–382.

LEWIS, M. Q. Short-term memory items in repeated recall. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 190–193.

RUNDUS, D., & ATKINSON, R. C. Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 99–105.

(Received October 28, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 17–24

RETENTION INTERVAL HYPOTHESIS AND EVIDENCE FOR ITS BASIC ASSUMPTIONS[1]

CHIZUKO IZAWA[2]

State University of New York at Buffalo

A new hypothesis is proposed to account for both superior and equal performances under the study-test (RT) method as compared with the anticipation method; up to now comparisons have been equivocal. The retention interval hypothesis, based on statistical processes of retention (R-T) interval distributions, assumes no differential acquisition processes but different retention processes between the methods, derived from temporal analyses of R (reinforcement)-R, T (test)-T, T-R, and R-T intervals defined in terms of individual items in paired-associate and verbal discrimination learning. Short-term retention losses occurring in R-T intervals in acquisition significantly controlled performance levels. Given the same mean R-T intervals, performances were the same in both short-term (Exp. I) and long-term (Exp. II) retention. In contrast, effects of feedback delay and task alternations were negligible. The results provided substantial support for the hypothesis and its basic assumptions.

In verbal learning, the anticipation method has often been compared with the study-test (reinforcement-test, RT) method. Yet, results were generally regarded as ambiguous or inconsistent, since there are reports claiming both superiority of RT over anticipation (e.g., Battig & Brackett, 1961; Cavanaugh & Parkman, 1971; Cofer, Diamond, Olsen, Stein, & Walker, 1967, Exp. I–II; Goss & Nodine, 1965, Exp. IX) and no differences (e.g., Bruder, 1969; Cofer et al., 1967, Exp. III–VI; Goss & Nodine, 1965, Exp. VII–VIII, X–XI; Lockhead, 1962). Both empirical and theoretical inconclusiveness might have come from the fact that previously the emphasis was upon the acquisition phase: Which method produces better learning? An alternative approach was taken here with the primary focus on the retention phase. What is acquired on an R trial (reinforcement, a presentation of both stimulus and response terms of a pair) must be retained at least until its subsequent T trial (test, a presentation of the stimulus term alone). Hence, retention processes, as well as acquisition processes, may control performance differences.

To investigate this possibility in paired-associate learning (PAL), procedural differences were examined in terms of time intervals based on individual items by specifying their presentation histories. Since an item is shown as either an R or a T on each trial, four exhaustive parameters result: R-R (inter-R-trial), T-T (inter-T-trial), T-R (feedback), and R-T (retention) intervals. The length of each may be determined by the number of its intervening events (both Rs and Ts) in an n-pair list, as exposure rates and intercycle intervals are, respectively, held constant under both anticipation and RT methods. The mean intervening Rs and Ts are the same, one half each for each time interval.

The R-R interval, the period from an R to the next R for the same item, directly controls acquisition based on R-spacing effects. With a random presentation, the distributions of n R-R intervals under anticipation range from 1 to $4n - 3$ intervening events with the mean of $2n - 1$;

---

[1] The present study was supported in part by Grants 050-7146-B and -C and 050-89604 of the State University of New York Research Foundation under a National Institutes of Health institutional research grant. Portions of the research were presented at the eleventh meeting of the Psychonomic Society in St. Louis, November 1971, and at the Twentieth International Congress of Psychology in Tokyo, August 1972. The author is grateful to D. D. Wickens, N. J. Slamecka, and B. R. Bugelski for their encouragement and to Christine Phytel (Exp. I) and Fred Sacks (Exp. II) for their competent assistance as *E*s.

[2] Requests for reprints should be sent to Chizuko Izawa, who is now at the Department of Psychology, Tulane University, New Orleans, Louisiana 70118.

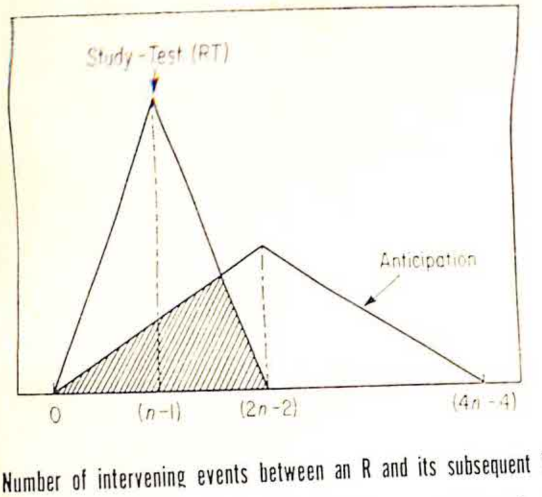


FIG. 1. Theoretical distribution curves of n individual retention (R-T) intervals within the list presented in a random order under both anticipation and RT (study-test) methods.

whereas under RT, they range from n to $3n - 2$ with the mean of $2n - 1$. The same boundaries and means hold, respectively, for T-T intervals. The facts suggest that, on the average, basic acquisition (R-R) and recalling task (T-T) processes may not differ between the methods. A frequently noted factor is the feedback interval, which is always 0 under anticipation and varies from 0 to $2n - 2$ with a mean of $n - 1$ events under RT. If feedback delay were really detrimental, the RT method should be inferior, which has not been borne out, however.

A critical difference between the two methods was uncovered to be the retention (R-T) interval, the period between an R and its subsequent T. Under anticipation the shortest R-T interval occurs when a given item is presented last on one anticipation cycle (one run-through of the n items) and first on the next, with 0 intervening event. The longest one occurs when an item is first on one cycle and last on the next, with $4n - 4$ events. The other cases vary somewhere between these two boundaries, with a mean of $2n - 2$. A similar analysis shows the retention intervals under RT vary from 0 to $2n - 2$ with the mean of $n - 1$ events. The mean absolute R-T intervals for anticipation and RT may be expressed, respectively, as: $2(n - 1)d +$ (intercycle interval) and $(n - 1)d +$ (intercycle interval). Both d (exposure duration) and intercycle intervals are, respectively, identical under both methods. Thus, the crucial difference boils down to different numbers of intervening events in R-T intervals: anticipation suffering from twice as many intervening events as RT, the difference being as large as $n - 1$ intervening events.

The longer R-T interval filled with interference producing events may constitute a critical disadvantage for anticipation, allowing more short-term losses than RT. This position is abundantly supported in data (e.g., Izawa, 1967; Murdock, 1961; Runquist, 1965; Tulving & Arbuckle, 1963). Accordingly, the retention interval hypothesis is proposed to account for procedural differences in PAL: *Other things equal, the RT method with the shorter mean retention interval with fewer intervening events should produce relatively better performance in acquisition than the anticipation method with the longer interval having more intervening events.*

To elaborate the hypothesis further, the distributions of n R-T intervals under a random item presentation are specified. Their curves as a function of intervening events are triangular as shown in Fig. 1 for both methods. The closer an item is to the zero point, the less the short-term forgetting, thus the better the performance: The relative advantage of RT over anticipation is evident. However, it is of supreme importance to note the relatively large overlapping area under both curves, indicating the existence of some probability that the two methods may produce identical performances. Therefore, the retention interval hypothesis can predict both superiority and equality of the RT method in contrast to the anticipation method. Hence, the results, previously interpreted as contradictory or inconsistent, may now be logically explained in the light of this hypothesis.

In addition to unmixed-list PAL with a random item presentation, the retention interval hypothesis holds equally well both theoretically (R-R, T-T, T-R, and R-T analyses) and empirically in verbal discrimination learning (VDL), constant presentation orders, mixed lists, part–whole presentations, and backward association learning. However, when the core time relations specified for the hypothesis do not hold, neither does the hypothesis: The outside domain includes probability and serial learning—the former due to different natures of tasks under the methods and the latter due to deviant temporal relationships.

Two experiments were conducted here to test some of the basic assumptions of the hypothesis applicable to various PAL and VDL situations.

EXPERIMENT I

Major theoretical positions taken previously were based on feedback delay and task alternations. Experiment I (four conditions) was designed to investigate effects of retention intervals of the present hypothesis as compared with those of critical factors of the alternative positions.

From temporal analyses above, despite identical mean R- and T-spacing effects, RT and anticipation methods differ sharply in feedback (T-R) and retention (R-T) intervals. Their effects may produce opposite outcomes: From feedback, anticipation with no delay should be superior to RT with the mean delay of $n - 1$ events. From the present hypothesis, however, RT with the shorter mean R-T interval ($n - 1$) is expected to perform better than anticipation with the longer one ($2n - 2$). To determine which factor controls PAL more predominantly, a situation was set up within a list design in such a way that predictions could be made in one direction from the feedback position and in the other from the retention interval hypothesis.[3]

Method.—A list of 20 CVC two-digit pairs taken from Izawa (1967) was divided into two equally difficult sublists, X and Y, of 10 pairs each. They were learned by 48 psychology students at State University of New York at Buffalo, 12 in each condition assigned according to the order of their appearance. Exposures of each R and T were 3 sec. under both methods with 3-sec. intercycle intervals. The 10 items within each sublist were randomized from cycle to cycle.

Under the RT method, Cond. 1 and 2 were designed holding both mean R-R and T-T intervals constant for both sublists but varying their mean retention and feedback intervals simultaneously by locating X and Y as follows:

RT Cycles:	T_0	R_1	T_1	R_2	T_2 ··· R_{10}	T_{10}
Cond. 1:	X Y	Y X	X Y	Y X	X Y···Y X	X Y
Cond. 2:	X Y	X Y	X Y	X Y	X Y···X Y	X Y.

In Cond. 1, the pairs of Part Y were shown always during the first half on all R cycles and during the last half on all T cycles: Their R-T intervals were distributed from n to $2n - 2$ with a mean of $3n/2 - 1$ events; whereas T-R intervals were from 0 to $n - 2$ with a mean of $n/2 - 1$. The Part X items resulted in the opposite relationships: a mean R-T interval of $n/2 - 1$ and a mean T-R interval of $3n/2 - 1$. Thus, Cond. 1 provided short feedback and long retention intervals for Part Y, and long T-R and short R-T for Part X, contrasting both factors within the RT method.

To verify the hypothesis differently, each sublist in Cond. 2 was arranged so that each had the same R-T interval, ranging from $n/2$ to $3n/2 - 2$ with a mean of $n - 1$, predicting the same performance for both Parts X and Y.

Under the anticipation method, feedback intervals cannot be varied but retention intervals can be, as done for Cond. 3 and 4:

Anticipation Cycles:	1	2	3	4 ··· 10	11
(Preceding Rs)	(0)	(1)	(2)	(3) ··· (9)	(10)
Cond. 3:	X Y	X Y	X Y	X Y···X Y	X Y
Cond. 4:	X Y	Y X	X Y	Y X···Y X	X Y.

In Cond. 3, in addition to the same mean R-R and T-T intervals, R-T intervals were the same for both X and Y, ranging from n to $3n - 4$ with a mean of $2n - 2$.

In manipulating different R-T intervals under anticipation, it is impossible to create uniformly long or short mean intervals through all acquisition stages. Nonetheless, Cond. 4 was developed to allow orderly variations: R-Ts for Part X were long from Anticipation Cycle 1 to 2, 3 to 4, . . . , 9 to 10; and

---

[3] The exact R-T intervals could easily be achieved for selected items only by dummies as done in short-term memory designs. However, the current aim is to understand differential performance between anticipation and RT under traditional list designs considering *all* items without interference from external items. Therefore, the required manipulations were achieved within the same traditional framework.

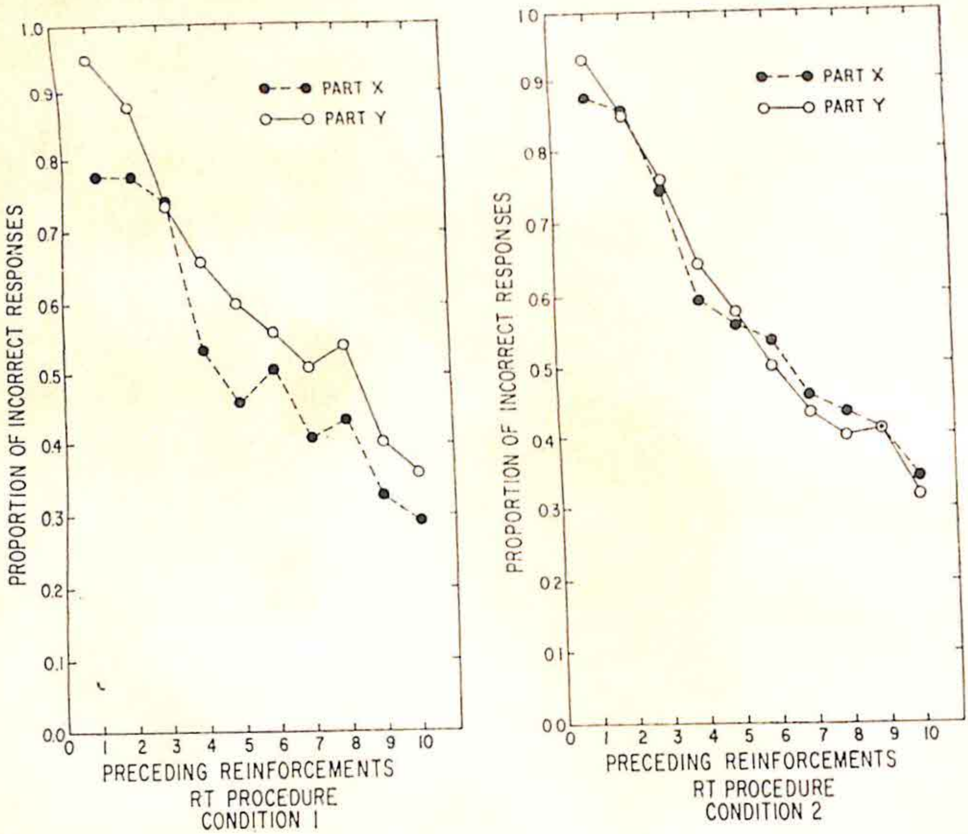


FIG. 2. Performances of Sublists X and Y in Cond. 1 (experimental) and 2 (control) under the RT method in Exp. I, as a function of the number of reinforcements (Rs) preceding each test (T).

short from Cycle 2 to 3, 4 to 5, . . . , 10 to 11. The reverse held for Part Y: R-Ts were short from each odd-numbered cycle to the next, and long from each even cycle to the next. The short R-T intervals distributed from 0 to $2n - 4$ with a mean of $n - 2$ events, and the long ones from $2n$ to $4n - 4$ with a mean of $3n - 2$. The T-R intervals for both X and Y were identical. No *S*s in Exp. I detected presentation structures.

Results and discussion.—The left panel of Fig. 2 gives data of Parts X and Y in Cond. 1 (RT) with respect to probability of incorrect responses as a function of the number of R trials experienced prior to a given test. Differences in total errors/*S* between sublists were large: $t(11) = 2.203$, $p < .05$. Significant superiority of Part X with shorter R-T intervals ($n/2 - 1$) over Part Y with longer ones ($3n/2 - 1$) provided support for the retention interval hypothesis. Evidently, the counteracting effects of feedback delay (T-R), if any, were overpowered by the effects of R-T intervals.

When R-T intervals were equated as in Cond. 2 (RT), both X and Y produced the same performances (Fig. 2, right) in the direction predicted from the hypothesis. Although the identity could also be expected from the feedback delay, its effects might be minor as seen in Cond. 1. When the two RT conditions were compared, nearly all points of Cond. 2 ($n - 1$ R-T events) fell between the Cond. 1 curves in agreement with the hypothesis: above X (shorter, $n/2 - 1$) and below Y (longer, $3n/2 - 1$).

Data in Cond. 3 were very similar to those in Cond. 2. Here also, under anticipation when R-T intervals were the same, no differences were observed between Parts X and Y, thus supporting the hypothesis.

If, in Cond. 4, feedback interval (always 0) had been a controlling factor, no differences should have occurred either between Parts X and Y or between adjacent cycles within each half list. None of these expectations, however, was borne out, since significant differences resulted in relevant comparisons. Thus, the feedback delay did not seem to produce effects of appreciable magnitude under the anticipation method either. In contrast, if the retention interval hypothesis is correct, the following specific predictions can be made and tested (Fig. 3):

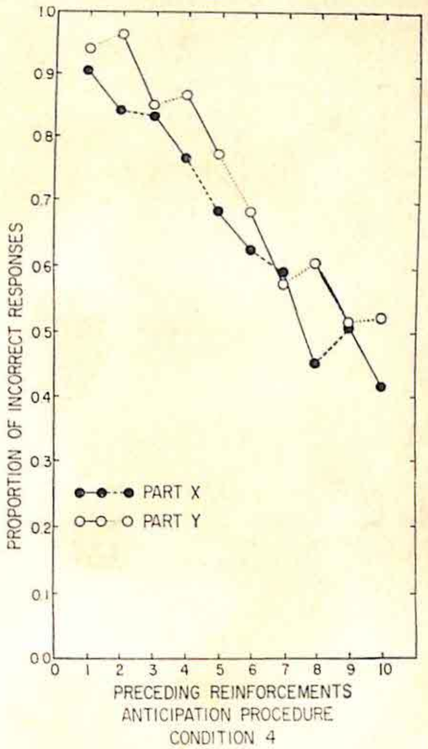


FIG. 3. Performances of Sublists X and Y in Cond. 4 (experimental) under the anticipation method in Exp. I. (The long R-T intervals are shown in dotted lines, and the shorter ones in solid lines.)

1. In accordance with alternate short and long R-T intervals, the X curve should show steeper progress from every odd-numbered R (even-numbered anticipation cycle) to the next with a mean of $n - 2$ events (short) and slower progress from each even R to the next with a mean of $3n - 2$ (long). Similarly, faster and slower progress should alternate, starting from the slower one (R_1) for Part Y. These theoretical predictions were fairly well borne out as seen in Fig. 3. Cycles with longer R-T's (dotted lines) were inferior despite their counteracting advantageous R-R and T-T intervals (Izawa, 1971a, 1971b); whereas those with shorter R-Ts were superior despite their disadvantageous R-R and T-T intervals. At times short-term losses overpowered the gain in acquisition. Differences between short and long R-Ts were substantial: $t(11) = 2.54$, $p < .05$ for X; and $t(11) = 1.97$, $.05 < p < .10$ for Y.

2. Since faster and slower progress rates for each sublist came alternately on different cycles, it is predicted that differences between Parts X and Y within the same cycle should be small following each odd-numbered R, but large following every even R. The prediction was strongly supported. Differences between the odd and even cycles were highly significant: $t(11) = 6.20$, $p < .001$.

Conditions 1 (RT) and 4 (anticipation) were manipulated with an additional purpose. Another explanation often noted is the effect of task alternation, stating that anticipation where R and T alternate constantly should be inferior to RT where R and T tasks are separated on independent cycles. Since Sublists X and Y were administered within the same procedure, alternation effects were necessarily identical for both X and Y in each condition. Therefore, the position with no specification of R-T intervals cannot predict differential performance between the sublists. Nevertheless, significant differences did occur in both cases, indicating that effects of task alternations may not be as influential as those of R-T intervals. Another evidence of this weakness was reported in Cofer et al. (1967, Exp. VI) where no improvement was rendered even when task alternation frequencies were reduced into a half. A further difficulty: The task alternation position may handle the superiority of RT, but it cannot accommodate the equality of the two methods, which is also reported in the literature.

To facilitate further theoretical considerations, Parts X and Y were combined for each condition in Fig. 4. For both controls, Cond. 2 (RT) and 3 (anticipation), the combined curves were practically identical to the separate ones as Parts X and Y. This was expected from the hypo-

6·3·81

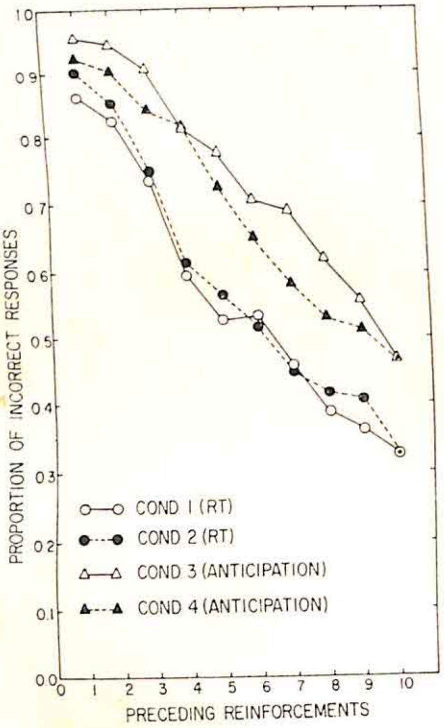


FIG. 4. Performance curves when both Sublists X and Y were combined ($X + Y = n$, the whole list) in all conditions in Exp. I.

thesis: Each mean R-T interval for the combined curves was identical to that for each sublist: $n - 1$ events for RT and $2n - 2$ for anticipation. For experimental groups, when significantly different X and Y under RT (Cond. 1) were combined, the resultant curve did not differ from its combined control. Similarly, complex but orderly differences between the sublists under anticipation (Cond. 4) disappeared when both were summed, revealing no differences from its control.[4] The identity between the combined experimental and control curves within each method is also in agreement with the hypothesis: The mean R-T intervals of $X + Y$ curves of both experimental and control groups were identical under each method.

Since no difference existed between either two RT or two anticipation conditions, the two respective conditions were further combined and procedural differences were tested. RT with a shorter mean retention interval ($n - 1$) significantly excelled anticipation with a longer one ($2n - 2$): $t(46) = 3.09$, $p < .01$, in support of the retention interval hypothesis.

Thus, whereas effects of feedback and task alternations were negligible, Exp. I results demonstrated that the R-T interval in acquisition seems to be a major determinant of performance differentials, if any, under both anticipation and RT methods, supporting the hypothesis.

EXPERIMENT II

There is both theoretical and empirical evidence that major differences between anticipation and RT methods may arise from short-term losses in R-T intervals but not from acquisition processes per se. To pursue this further in Exp. II, after acquisition was completed, for both methods retention was measured for the same retention period which was much longer than usual acquisition R-T intervals.

Method.—Two lists of 12 CVC-CVC intermediately difficult pairs were learned at a 2-sec. rate with 2-sec. intercycle intervals under anticipation (Cond. 1 & 3) and RT (Cond. 2 & 4) with 10 State University of New York at Buffalo *S*s in each condition, the total being 40. To equate the total Rs for both methods, Ts only were given for the final anticipation cycle by replacing all Rs by their Ts (dummies), thus producing the same R-T intervals on the last cycle as on all others for the valid Ts. Although anticipation had two successive Ts after R_{10}, whereas RT had one T, it was unlikely that this treatment distorted data, since an additional T *not* followed by an R produced no significant effects on subsequent performance (Izawa, 1968, Cond. RT vs. RTT). After acquisition, a 15-min. ability test (neutral to PAL) was interpolated, followed by two retention Ts. The 15-min. (900 sec.) retention period was quite long as compared with mean acquisition R-T intervals (24 sec. for RT and 48 sec. for anticipation).[5]

Results and discussion.—Data for Lists 1 (left) and 2 (right) are shown in Fig. 5.

---

[4] This state of affairs may depict the fact that some aspects of the processes may never be detected when the mean performances of the entire list are the sole reference point of theoretical and empirical considerations. Here is another reason why analyses in terms of individual items, as performed in the present hypothesis, may have significance.

[5] In a strict sense, the mean intervals between the last R and the first retention T were 924 sec. under RT and 948 sec. under anticipation. However, the difference of 24 sec. here can be regarded as negligible in long-term memory.

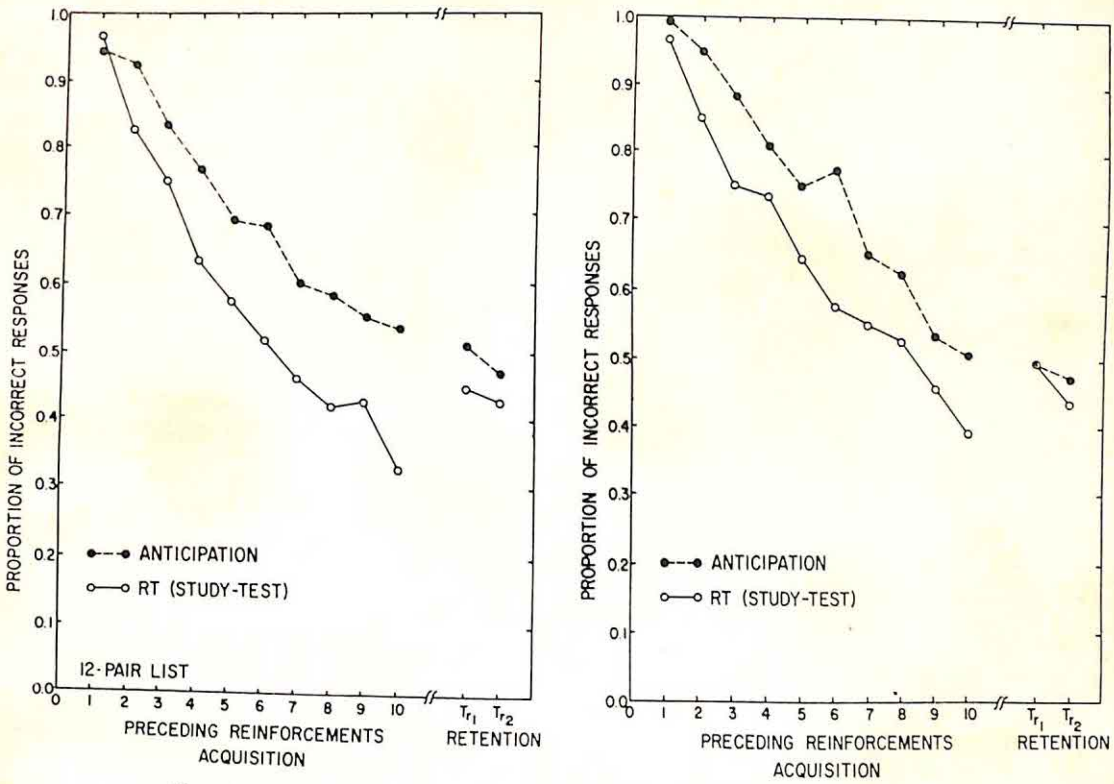


FIG. 5. Acquisition and retention data for Cond. 1 and 2 (left, List 1) and Cond. 3 and 4 (right, List 2) in Exp. II.

In both cases, the hypothesis was supported since acquisition performances were significantly better for RT than anticipation ($p < .025$). After the interpolated task, however, an entirely different picture emerged: The large differences during acquisition virtually disappeared. Given approximately the same retention intervals, no differences resulted between the two methods in long-term memory.

To explain these results, it is unconvincing to appeal to differential rehearsal during 15 min., since the majority of *S*s reported being "too busy" with the ability test "to worry about anything else" regardless of the conditions. Another possible interpretation might be: Learning was greater but retention was poorer under RT, and the reverse held for anticipation. Such a position must assume the following simultaneously: (*a*) two acquisition processes, the better one for RT and the poorer one for anticipation, and (*b*) two opposing retention processes within each method. Retention during acquisition (short-term) is better but suddenly changes to poorer during the retention period (long-term) only after a 15-min. neutral task for RT, whereas the reverse holds for anticipation. Both assumptions have difficulties: (*a*) It is incorrect to assume different acquisition processes, since both methods have been reported to produce equal learning. (*b*) It is difficult to build a logical theory which encompasses opposite processes between short- and long-term memories: When short-term memory is poor to begin with, as under anticipation, it is hard to assert that its long-term memory is better nonetheless, since abundant evidence indicates that short- and long-term memories are correlated. Further, this position requires theoretical mechanisms dealing with a complete shift during a 15-min. interval for both methods, which may be difficult to maintain.

In contrast, the present hypothesis requires only one concept of retention losses based on different retention intervals and can parsimoniously and logically explain major aspects of seemingly complex data: Given different R-T intervals as in present acquisition tests, the two methods produced significantly different performances. However, given approximately the same R-T intervals as in retention tests, the methods produced the same performances. These results are consistent with those in Cond. 2 and 3 in Exp. I where identical acquisition performances were obtained, provided there were identical mean R-T intervals. That is, both methods produced the same results, given the same retention intervals not only in short-term (acquisition period) but also in long-term (retention period) memory.

Over a 15-min. retention period forgetting was limited to RT, with no changes observed for anticipation (Fig. 5). Apparently short-term losses following each R occur fairly rapidly prior to the time $2(n-1)d$ + (intercycle interval), as in the last anticipation acquisition T, cf. $(n-1)d$ + (intercycle interval) under RT, and reach an asymptote soon thereafter, producing no further losses.

Findings in both Exp. I and II confirmed theoretical deductions from the hypothesis based on temporal parameter analyses: Both anticipation and RT have the same mean inter-R-trial (R-R) intervals which control major acquisition processes, suggesting similar acquisition for both. Yet the methods sharply differ in retention (R-T) intervals which control major retention processes, suggesting different retention processes. Both deductions were empirically verified, supporting the basic assumptions underlying the proposed hypothesis.

Thus, the function of retention intervals suggests itself as a major factor, both theoretically and empirically, for differentiating anticipation and RT (study-test) methods. The retention interval hypothesis logically explains the procedural differences in performance previously interpreted as inconsistent, ambiguous, and puzzling by encompassing both identity and superiority of RT over anticipation in PAL and VDL.

REFERENCES

Battig, W. F., & Brackett, H. R. Comparison of anticipation and recall methods in paired-associate learning. *Psychological Reports*, 1961, **9**, 59–65.

Bruder, G. A. Effects of method of presentation on paired-associate learning. *Journal of Experimental Psychology*, 1969, **79**, 383–384.

Cavanaugh, J. P., & Parkman, J. A comparison of the study-recall and anticipation methods in steady state paired-associate learning. *Psychonomic Science*, 1971, **22**, 361–363.

Cofer, C. N., Diamond, F., Olsen, R. A., Stein, J. S., & Walker, H. Comparison of anticipation and recall methods in paired-associate learning. *Journal of Experimental Psychology*, 1967, **75**, 545–558.

Goss, A. E., & Nodine, C. F. *Paired-associate learning*. New York: Academic Press, 1965.

Izawa, C. Function of test trials in paired-associate learning. *Journal of Experimental Psychology*, 1967, **75**, 194–209.

Izawa, C. Effects of reinforcement, neutral and test trials upon paired-associate acquisition and retention. *Psychological Reports*, 1968, **23**, 947–959.

Izawa, C. Massed and spaced practice in paired-associate learning: List versus item distributions. *Journal of Experimental Psychology*, 1971, **89**, 10–21. (a)

Izawa, C. The test trial potentiating model. *Journal of Mathematical Psychology*, 1971, **8**, 220–224. (b)

Lockhead, G. R. Methods of presenting paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 62–65.

Murdock, B. B., Jr. The retention of individual items. *Journal of Experimental Psychology*, 1961, **62**, 618–625.

Runquist, W. N. Intertrial oscillation and intertrial forgetting in paired-associate learning. *Canadian Journal of Psychology*, 1965, **19**, 145–154.

Tulving, E., & Arbuckle, T. Y. Sources of intratrial interferences in immediate recall of paired-associates. *Journal of Verbal Learning and Verbal Behavior*, 1963, **1**, 321–334.

(Received December 16, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 25–31

ORIENTATION ILLUSION AND MASKING IN CENTRAL AND PERIPHERAL VISION[1]

RAY OVER,[2] JACK BROERSE, AND BORIS CRASSINI

University of Queensland, St. Lucia, Brisbane, Queensland, Australia

A vertical line appears tilted when viewed following inspection of tilted lines (aftereffect) or against a background of tilted lines (illusion). In Exp. I the illusion obtained, as the orientation of the inducing lines was varied, differed between central and peripheral vision. Related differences have earlier been reported for the tilt aftereffect, and were interpreted as indicating that orientation-selective cells in the human visual system are more broadly tuned in peripheral than central vision. This possibility was examined in Exp. II by use of a contour masking paradigm. The detectability of a vertical line varied with the orientation of a grating exposed immediately before the line, but contrary to expectations, masking (threshold elevation) was obtained over a wider range of orientation values in central than peripheral vision. No evidence was obtained that perceptual facilitation (improved detection of the line following exposure to the grating) can occur in central vision in addition to masking. The present results question whether aftereffect, illusion, and masking paradigms are equally effective in obtaining information about the tuning characteristics of feature analyzers in the human visual system.

A vertical test line looks slightly tilted immediately after inspection of tilted lines (aftereffect) or when viewed against a background of tilted lines (illusion). In central vision the vertical line appears tilted clockwise following inspection of lines tilted less than 45° in the counterclockwise direction from vertical (angular repulsion). Angular attraction occurs when the difference in tilt between the inspection and test contours is greater than 45° (Gibson & Radner, 1937). In peripheral vision, angular repulsion is found at every tilt value of the inspection figure to 90° (Muir & Over, 1970).

Recent explanations of tilt aftereffects (Coltheart, 1971; Day, 1969; Over, 1971) have attributed angular distortion found when one line is viewed following inspection of another line to inhibitory interaction between neural analyzers which signal the orientation of each line. Over has suggested that the difference in angular function between visual tilt aftereffect in central and peripheral vision reflects differences in the breadth of tuning of tilt analyzers at the two retinal locations. He proposed that in central vision each analyzer is maximally excited at a specific contour orientation, less excited as the line's tilt is varied over 30° or so, and silent (relative to the unit's activity during homogeneous stimulation) when the stimulating line is displayed at, or nearly at, right angles to the preferred direction. In these terms angular repulsion occurs because cells tuned to both the test and inspection orientations are suppressed when the inspection figure is replaced by the test figure. Angular attraction is attributed to postexcitatory facilitation, with cells silent during inspection being overexcited rather than suppressed when the test line is introduced. In explaining the absence of angular attraction in peripheral vision as well as the extensive angular range over which repulsion occurs, Muir and Over (1970) and Over proposed that orientation analyzers in peripheral vision have wider excitatory ranges than units in central vision.

In Exp. I, the intersect illusion is measured in central and peripheral vision as a function of the orientation of the inducing lines against which the test line is seen.

---

[1] This research was supported by a grant to the first author from the Australian Research Grants Committee. Thanks are due to Ann-Marie Parker for assistance in running the first experiment.

[2] Requests for reprints should be sent to Ray Over, Department of Psychology, University of Queensland, St. Lucia, Brisbane, Queensland, Australia 4067.

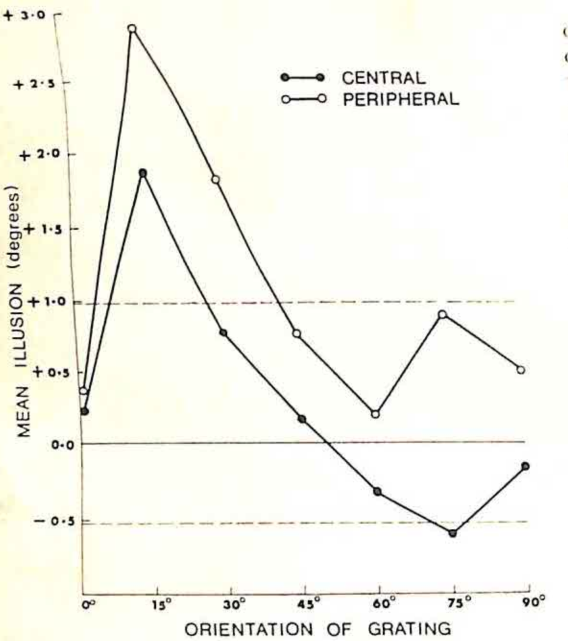


FIG. 1. Mean illusion in central and peripheral vision for the different orientations of the inspection grating.

There have been recent claims (e.g., Coltheart, 1971; Over, 1971) that the same error-inducing mechanism is responsible for orientation aftereffects and illusions. If this is the case, the intersect illusion should be exclusively of the contour repulsion form in peripheral vision, but in central vision should change from contour repulsion to attraction as the inducing and test lines differ in orientation by more than 45°. The illusion has not previously been studied in peripheral vision. In central vision, Gibson (1937a) and Wallace and Crampton (1969) obtained angular repulsion up to 45°, but did not measure the illusion when the test and background lines were separated by more than 45°. Logan (1962) demonstrated a transition from repulsion to attraction, but Blakemore, Carpenter, and Georgeson (1970) found no evidence of contour attraction when the test and inducing lines differed by between 45° and 90°.

EXPERIMENT I

Method

Subjects.—Sixteen undergraduate students with normal or corrected-to-normal vision served as *Ss*.

Apparatus.—A single test line was superimposed optically on a circular square-wave grating by use of separate channels of a Gerbrands three-field tachistoscope (Model T-3B-1). The periodicity of the grating, which subtended 2° 40′, was 4.2 cycles/degree, and the contrast was .89. In the course of the experiment the grating was tilted 0°, 15°, 30°, 45°, 60°, 75°, or 90° from vertical, and a homogeneous field of the same luminance (6.86 cd/m^2) as the space-average luminance of the grating was used as an inducing figure to obtain control measures. The test line was a length of white tape, 2° 15′ × 7′, fixed on black cardboard.

Procedure.—The *S*'s task was to judge the orientation of the test line when it was displayed in either central vision (eight *Ss*) or in peripheral vision 7° to the right of the fixation point (eight *Ss*). Under both conditions, the fixation point was shown for 1 sec. in the third channel of the tachistoscope before the test and inducing figures were presented. Each *S* judged the tilt of the test line with the grating displayed at each of the seven orientations as well as with the homogeneous background, and the order in which the eight conditions were tested was varied across *Ss* in each group by a Latin square design. With each induction condition, the test line was initially shown at the vertical position for 15 msec. and either the grating or the homogeneous field was simultaneously presented as the background. The *S* used a hand switch to initiate trials on *E*'s command, and trials were separated by 20 sec. On every trial, *S* was required to report whether the test line appeared tilted to the left or to the right of vertical; judgments of vertical were not permitted. After each trial *E* varied the orientation of the test line by 1° in the direction opposite to the tilt last reported by *S*. This procedure continued until *S* reversed successive judgments; then *E* shifted the test line by 1° in either the clockwise or counterclockwise direction, determined at random. This procedure continued until six reversals in judgment had occurred under the specific condition; further measures were then obtained using different background conditions.

Results

The position midway between tilt values on which *S* reversed successive judgments was taken as the orientation at which the test line appeared vertical to *S*. Measures obtained with conjoint presentation of the test line and the grating were subtracted from measures found using the homogeneous field as background to establish the magnitude of the illusion. The values in Fig. 1 show the mean shift in judgment induced in central and peripheral vision by display of the grating at the seven orientations which were used. The symbol + is used to signify that the apparent vertical was shifted in the direction of tilt of the grating, and − that the distortion was in the opposite direction.

An analysis of variance based on measures obtained at grating orientations between 15° and 90° revealed that mean illusion varied as a function of the angle of tilt of the grating, F (5, 70) = 32.00, $p < .01$, and differed between central and peripheral vision, F (1, 14) = 13.86, $p < .01$. The interaction between angle and viewing condition was not significant, F (5, 70) = 1.34, $p > .05$. Comparisons using Duncan's new multiple-range test indicated that the mean illusion in peripheral vision was significantly larger ($p < .05$) than the corresponding error in central vision when the grating had been tilted 15, 30, or 75°. Figure 1 shows by dotted lines the confidence limits ($p < .05$) when the illusion value obtained with the 0° inducing grating in central vision served as a base line against which other values were compared by Duncan's test.

Discussion

The difference in angular function for the intersect illusion between central and peripheral vision parallels results obtained with the tilt aftereffect (Muir & Over, 1970). In each case an asymmetrical S-shaped angular function is obtained when the test line and grating are fixated; perceptual repulsion results when the figures differ in orientation by less than 45° and contour attraction is found when the tilt difference is more extreme. In peripheral vision perceptual repulsion can alone be induced. The illusion measures obtained in the present experiment are considerably larger in absolute terms than the aftereffect values reported by Muir and Over (1970), but it is probable this difference simply reflects the procedural artifact by which decay occurs in aftereffect in the period between cessation of inspection and completion of the test judgment.

The results of Exp. I suggest that the same error-inducing mechanism underlies the tilt aftereffect and illusion, but they do not permit direct identification of the mechanism. The claim (e.g. Over, 1971) that differences in angular function for the aftereffect and illusion between central and peripheral vision reflect differences in the breadth of tuning of neural analyzers at the two retinal locations can be examined further only if information is available about orientation tuning in the human visual system. Experiment II studied contour masking in central and peripheral vision for this reason.

EXPERIMENT II

Weisstein (1969) has argued that the response properties of neural analyzers in the human visual system can be established using psychophysical masking paradigms. In contour masking studies *S*s are required to detect a single line or grating, the target stimulus, displayed briefly before (Sekuler, 1965), together with (Campbell & Kulikowski, 1966), or immediately following (Houlihan & Sekuler, 1968) exposure to another line or grating, the masking stimulus, in the same part of the visual field. With each arrangement the detectability of the target is most impaired when the masking and test lines are identical in orientation, and improves as their tilts diverge. Masking has been attributed to inhibitory interaction between neural units with overlapping sensitivities, and the extent to which a line at one tilt masks a line at another tilt is considered an index of the tuning of orientation analyzers in the human visual system (see Weisstein, 1969). Masking measures must, however, reflect the pooled response of many neural analyzers. Orientational selectivity has been demonstrated directly by microelectrode study of single cells in the visual cortex of the cat (Campbell, Cleland, Cooper, & Enroth-Cugell, 1968) and the monkey (Hubel & Wiesel, 1968). These cells are selective to the orientation of edges but are relatively unresponsive to diffuse light. Each cell discharges maximally to a particular contour orientation, and is progressively less excited over an angular range (20° to 30°) on either side of the preferred direction. Microelectrode techniques so far have not been explicitly used to compare the breadth of tuning of orientation-selective cells in central and peripheral vision.

In Exp. II contour masking is studied in central and peripheral vision as the tilt between the masking and test stimuli is increased from 0° to 90°. The issues of interest were whether masking occurs over a greater angular range in peripheral than central vision and whether facilitation (as opposed to masking) is found in central vision when the inducing and test lines differ in orientation by 70° to 90°. In the latter connection Gilinsky (1967, 1968) found that detection of a vertical line displayed following exposure to a horizontal grating improved as the duration of the mask was increased. Woll, Eriksen, and Hake (1970) have criticized the experimental procedures adopted by Gilinsky; using a forced-choice methodology they obtained no evidence of facilitation when the inducing and test stimuli differed in orientation by 90°. However the issue of whether facilitation occurs in addition to masking should be further examined by comparing detection measures following exposure to a grating with values obtained when a homogeneous field at the same space-average luminance as the grating is used as the inducing stimulus. This control has been employed in three previous studies of contour masking. With forward masking, Houlihan and Sekuler (1968) found that inspection of a 42° grating and a homogeneous field had similar influence on detection of a vertical line; measures were not obtained at more extreme tilts. With backward masking, Sekuler (1965) found that exposure to a 60° grating impaired detection of a vertical line, and only one of three *S*s tested by Parlee (1969) showed evidence of facilitation when mask and test lines differed by 75°. In the present study, the detectability of a vertical target line was measured as the orientation of a masking grating was varied from vertical to horizontal in 10° steps, and with a homogeneous inducing field used as a control condition.

Method

Subjects.—The eight *S*s included the three *E*s. Each *S* had normal or corrected vision.

Apparatus.—An adaptation field, the inducing stimulus, and the target stimulus were displayed successively in separate channels of a Gerbrands tachistoscope. A forward masking paradigm was used, with the adaptation field presented for 1,000 msec., the inducing field for 120 msec., and the target (or a homogeneous test field on 'catch' trials) for 10 msec. There was a dark interval of 120 msec. between offset of the adaptation field and onset of the inducing field, and of 10 msec. between offset of the inducing field and onset of the target.

Stimuli were constructed by fixing mat-black tape to white cardboard. The target was a single vertical line, 1°42′ long and 6′ wide; the grating used as the inducing stimulus subtended 60°3′ × 6°54′, and consisted of 12′ wide lines separated by 12′. The line and the grating were each displayed on a background 12°48′ × 14°20′. The Michelson contrast value of the grating was 1.0, and its space-average luminance was maintained at 4.35 cd/m^2. The space-average luminance of the target field was manipulated as the dependent variable of the experiment. Two fixation points, 7° apart, were present in each field. The fixation point to *S*'s left coincided spatially with the centers of the grating and the target line, and was used for central vision. For peripheral vision, *S*s viewed the right-hand fixation point which enabled the grating and the test line to be displayed 7° in the periphery. A dental impression was used to limit head movements during fixation.

Ten orientation values (extending from vertical to horizontal in 10° steps) of the grating were used. In addition, detection measures were obtained employing a homogeneous inducing field of the same space-average luminance as the grating.

Procedure.—In each session the threshold for detection of the target line was measured once at each orientation of the grating, as well as with the homogeneous inducing field. Each *S* was tested over four sessions, and judgments were made in central and peripheral vision on alternate sessions. Detection was measured using a block-wise tracking method adapted from Houlihan and Sekuler (1968). In each block there were six target and six catch trials presented in random order at a given luminance value. If for a given mask stimulus *S* was more than 75% accurate on the initial block, the luminance of the target was decreased over blocks in steps of 1.72 cd/m^2 until accuracy on the 12 trials of a block was 75% or less. The luminance value was increased over blocks in similar steps if *S* was initially less than 75% accurate. For each *S* the starting value for the central viewing track was determined by preliminary testing as the luminance of the target required for *S* to perform at chance level when the homogeneous inducing field was used as a mask. A similar procedure was used to determine a starting level in peripheral vision for each *S*.

The order of presentation of the inducing stimuli across *S*s and between sessions was varied by three 11 × 11 Latin squares, with a single sequence unused. The testing was conducted in a darkened room, and each *S* was dark-adapted for 10 min. prior to testing. There was a 10-sec. interval between successive trials. On each trial *S* initiated the stimulus display, and indicated his judgment to *E* by operation of a hand switch.

Results

The threshold for detection of the target line was defined as the luminance level at which *S* was able to differentiate the 12 target and catch trials in a single block with 75% accuracy. Where necessary the threshold measure was determined by extrapolation. Thresholds obtained with different orientations of the grating in central and peripheral vision were converted to ratio values relative to thresholds found in central and peripheral vision using the homogeneous inducing field. The latter values serve as base lines against which impairment in detection induced by the spatial properties of the grating can be established.

Because the group means and variances were correlated, a logarithmic transform was applied to the ratio values of each *S*. Figure 2 shows the mean log ratios obtained in central and peripheral vision at each angle of tilt of the grating. Masking is in evidence if the log ratio exceeds .0; facilitation (better detection of the target following exposure to the grating than to the homogeneous field) has occurred if the ratio is less than .0. The group data were then subjected to analysis of variance. As was expected from previous contour masking experiments, the relative detectability of the target was dependent on the orientation of the grating, $F\ (9,\ 63) = 134.03$, $p < .01$. Detectability did not differ significantly between central and peripheral vision, $F\ (1,\ 7) = 4.07$, $p > .05$. However, the interaction between retinal location and the orientation of the grating was significant, $F\ (9,\ 63) = 2.68$, $p < .02$. Comparisons made using Duncan's new multiple-range test showed that exposure to the grating induced significantly greater ($p < .01$) impairment in detection of the target line in central than peripheral vision at each grating orientation between 0° and 50°, except 40°. There was no significant difference ($p > .05$) in detection between central and peripheral vision when the target and the grating differed in orientation by more than 50°.

The confidence limit ($p < .05$), when the mean log ratio obtained in central vision with the grating oriented 90° is taken as a base line against which other values are compared by Duncan's test, is shown in Fig. 1 by dotted lines. In central vision exposure to the grating at any orientation between 0° and 50° resulted in impaired detection of the target relative to this base line, and facilitation did not occur at angular values between 50° and 90°. Using the same base line, masking in peripheral vision extended over only 30°, and significant facilitation (improved detection as a function of exposure to the grating) occurred at 80° but not at other values.

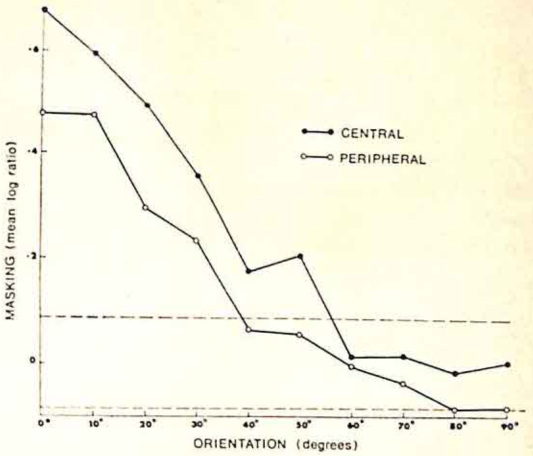


FIG. 2. Mean log. masking ratios in central and peripheral vision for the different orientations of the inducing grating.

DISCUSSION

It has recently been proposed (e.g. Blakemore et al., 1970; Coltheart, 1971; Over, 1971) that perceptual distortions found with simultaneous, as well as successive display of a grating and a test line, result from inhibitory interaction between the neural correlates of the two figures. The results of Exp. I indicate that fundamentally similar error-inducing processes are responsible for the visual tilt aftereffect and the intersect illusion. The mechanism of distortion remains, however, to be identified. Muir and Over (1970) and Over attributed differences in the angular function of the aftereffect between central and peripheral vision to variations in the breadth of tuning of orientation detectors within the visual field. An indirect test was made of this claim in Exp. II by examining orientation masking in central and peripheral vision. The results of Exp. II indicated that orientation masking and aftereffect are not equivalent perceptual phenomena. Tilt aftereffects are of greater magnitude in peripheral than central vision (Muir & Over, 1970), yet masking is more pronounced in central than peripheral vision. Similarly, tilt aftereffects can be generated over a wider range of angular values of inducing and test stimuli in peripheral than central vision, but the reverse is the case for orientation masking. In addition, the tilt aftereffect changes from proactive angular repulsion to attraction as the orientation difference between the inducing and the test stimuli is increased; there is no comparable change from impairment of detection to facilitation in the masking situation.

The different data functions between orientation masking and aftereffect are surprizing in view of procedural similarities between the traditional aftereffect paradigm and the cross-adaptation forward masking design. The procedures differ solely in that with masking the influence of the inducing stimulus on the threshold for detection of the target is measured, while with the aftereffect *S* is required to judge a specific property of the suprathreshold target following exposure to the inducing stimulus. Weisstein (1969) has argued that masking data alone provide information on specific feature-analyzing mechanisms, on the grounds that aftereffect measures reflect the pooled response of a number of different types of analyzers. Recent accounts of tilt aftereffects in terms of neural inhibition (Coltheart, 1971; Day, 1969; Over, 1971) merely suppose that perceptual distortion results because inspection induces selective response suppression within the population of tilt analyzers. If Weisstein's claim is correct, inferences could be drawn about tuning characteristics of specific feature detectors from aftereffect measures only if a model of aftereffect is developed that makes explicit reference to the manner in which the neural correlates of the inspection and test figures interact and also identifies the input properties on which judgments are based. The suprathreshold target in the aftereffect situation would undoubtedly have more extensive and massive neural representation than the barely detectable target in a masking study; judgments of the orientation of the target also may be based on different input characteristics than judgments of whether a target was present or absent. While these factors may suggest why masking occurs over a smaller range of angular values than aftereffect, they do not indicate why relative error in central and peripheral vision should be reversed for masking and aftereffect. Nor are clues provided as to the mechanism by which contour attraction, in addition to contour repulsion, occurs with tilt aftereffect and the associated illusion.

The relationship between masking and aftereffect has also been questioned by Sekuler, Lehr, Stone, and Wolf (1971), who demonstrated masking of perceived movement under conditions where the motion aftereffect cannot be induced. They found equivalent masking following exposure to downward moving lines and alternate presentation of upward and downward moving lines. The motion aftereffect can be generated by the former, but not by the latter, inspection conditions. However, Blakemore and Nachmias (1971) found that the extent to which exposure to one grating impairs detection (masking) or shifts the apparent periodicity (aftereffect) of another grating is dependent on the relative orientations of the two gratings as well as the extent to which they differ in spatial frequency. Orientation selectivity was identical for masking and aftereffect. The relationship between masking and aftereffect should receive further attention in view of these seemingly discrepant findings.

Last, the possibility exists that visual tilt aftereffects and illusions result from processes other than, or additional to, neural inhibition. Gibson (1937b), for example, proposed that tilt aftereffects occur as by-products of a psychophysical normalization process, and Morant and Harris (1965) have argued that both normalization and neural inhibition typically influence tilt aftereffects. Neither treatment is able, however, to account for existing data on aftereffects (see Over, 1971). It seems unlikely that the error-inducing mechanisms underlying tilt aftereffects and illusions will be identified until further variables affecting the magnitude of the aftereffect and illusion are experimentally differentiated from those that have no influence. If a feature-analyzer explanation is to be maintained, it will be necessary to show that the variables critical for generation of the aftereffect and illusion are also those to which neural analyzers tuned to contour orientation are differentially sensitive. In doing this it may be profitable to reexamine claims (e.g. Weisstein, 1969) that masking paradigms provide data on the tuning characteristics of feature detectors.

REFERENCES

Blakemore, C., Carpenter, R. H. S., & Georgeson, M. A. Lateral inhibition between orientation detectors in the human visual system. *Nature*, 1970, **228**, 37–39.

Blakemore, C., & Nachmias, J. The orientation specificity of two visual after-effects. *Journal of Physiology*, 1971, **213**, 157–174.

Campbell, F. W., Cleland, B. G., Cooper, G. F., & Enroth-Cugell, C. The angular selectivity of visual cortical cells to moving gratings. *Journal of Physiology*, 1968, **198**, 237–250.

Campbell, F. W., & Kulikowski, J. J. Orientational selectivity of the human visual system. *Journal of Physiology*, 1966, **187**, 437–445.

Coltheart, M. Visual feature-analyzers and aftereffects of tilt and curvature. *Psychological Review*, 1971, **78**, 114–121.

Day, R. H. *Human perception*. Sydney: Wiley, 1969.

Gibson, J. J. Adaptation, aftereffect, and contrast in the perception of tilted lines: II. Simultaneous contrast and the areal restriction of the aftereffect. *Journal of Experimental Psychology*, 1937, **20**, 453–467. (a)

Gibson, J. J. Adaptation with negative aftereffect. *Psychological Review*, 1937, **44**, 222–244. (b)

Gibson, J. J., & Radner, M. Adaptation, aftereffect, and contrast in the perception of tilted lines: I. Quantitative studies. *Journal of Experimental Psychology*, 1937, **20**, 453–467.

Gilinsky, A. S. Masking of contour-detectors in the human visual system. *Psychonomic Science*, 1967, **8**, 395–396.

Gilinsky, A. S. Orientation-specific effects of patterns of adapting light on visual acuity. *Journal of the Optical Society of America*, 1968, **58**, 13–18.

Houlihan, K., & Sekuler, R. W. Contour interactions in visual masking. *Journal of Experimental Psychology*, 1968, **77**, 281–285.

Hubel, D. H., & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 1968, **195**, 215–243.

Logan, J. A. An examination of the relationship between visual illusion and aftereffects. Unpublished doctoral dissertation, University of Sydney, Australia, 1962.

Morant, R. B., & Harris, J. R. Two different aftereffects of exposure to visual tilts. *American Journal of Psychology*, 1965, **78**, 218–226.

Muir, D., & Over, R. Tilt aftereffects in central and peripheral vision. *Journal of Experimental Psychology*, 1970, **85**, 165–170.

Over, R. Comparison of normalization theory and neural enhancement explanation of negative aftereffects. *Psychological Bulletin*, 1971, **75**, 225–243.

Parlee, M. B. Visual backward masking of a single line by a single line. *Vision Research*, 1969, **9**, 199–205.

Sekuler, R. W. Spatial and temporal determinants of visual backward masking. *Journal of Experimental Psychology*, 1965, **70**, 401–406.

Sekuler, R., Lehr, D., Stone, W., & Wolf, M. Human visual motion sensitivity: Evidence against a ratio theory of sensory coding. *Perception & Psychophysics*, 1971, **9**, 483–484.

Wallace, G. K., & Crampton, D. J. The effect of background density on the Zollner illusion. *Vision Research*, 1969, **9**, 167–177.

Weisstein, N. What the frog's eye tells the human brain: Single cell analyzers in the human visual system. *Psychological Bulletin*, 1969, **72**, 157–176.

Woll, S. Eriksen, C. W., & Hake, H. W. A forced-choice study of edge detectors in the human visual system. *Perception & Psychophysics*, 1970, **9**, 247–252.

(Received December 23, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 32–36

SEQUENTIAL EFFECTS IN SERIAL REACTION TIME[1]

NEIL H. KIRBY[2]

University of Adelaide, South Australia, Australia

The serial reaction time data from an experiment producing a negative recency effect at different interstimulus intervals were subjected to an analysis which included an investigation of up to and including fifth-order sequential effects. Third-, fourth-, and fifth-order effects were found to occur, indicating that the negative recency could not be explained by considering sequences of two events only. The implications of these findings are discussed with respect to models seeking to explain sequential effects in serial reaction time tasks.

Sequential effects have been found in serial reaction time experiments such that under some conditions, responses to a signal which is the same as the preceding signal are faster than if it is different (e.g., Bertelson, 1961). Under other conditions the opposite has been found, the response to a signal which is different from the preceding signal being faster than if it is the same (e.g., Williams, 1966). The former has been called a positive recency or "repetition" effect (Bertelson, 1961), and the latter a negative recency or "alternation" effect.

Some experiments have reported that the repetition effect tends to decrease, and even to change to an alternation effect, with increasing interstimulus interval (e.g., Bertelson, 1961; Hale, 1967), but others have failed to find such a change (e.g., Keele, 1969; Schvaneveldt & Chase, 1969). Generally, however, with two-choice compatible S–R arrangements, repetition effects have been found with interstimulus intervals of less than 1 sec. (e.g., Bertelson, 1961; Hale, 1967), while alternation effects have been found with interstimulus intervals of greater than 1 sec. (e.g., Hale, 1967; Williams, 1966).

Attempts to explain the repetition effect by Bertelson (1963) and the alternation effect by Williams (1966) have both assumed that they primarily are due to the immediately preceding event. However, Remington (1969) has shown that, for the repetition effect at least, events two, three, and four back in the series can also contribute to the difference between repetitions and alternations, and he has argued that these must therefore be taken into account in formulating any explanatory model.

The present experiment aimed at obtaining reaction time data showing alternation effects by using a two-choice compatible S–R arrangement with interstimulus intervals of greater than 1 sec., and at applying Remington's (1969) type of analysis in order to see whether higher order sequences also need to be taken into account in any explanation of the alternation effect.

Briefly Remington's (1969) analysis consists of separating out all the possible combinations of the two stimulus events in sequences of up to five in length. Thus the first-order stimulus reaction time, which may be represented by A, is the overall mean reaction time. The second order consists of 2 reaction times, that for the case when a stimulus is preceded by the same stimulus (AA) and that for the case in which it is preceded by a different stimulus (BA). Similarly there are 4 third-order reaction times corresponding to the four possible combinations, AAA, ABA, BBA and BAA; 8 fourth-order reaction times and 16 fifth-order.

METHOD

Subjects.—The *S*s were five male and one female volunteers who were naive as to the aims of the experiment and ranged in age from 20 to 40 yr.

---

[1] The author wishes to thank A. T. Welford for his valuable advice, encouragement, and criticism, and R. Willson for writing the computer programs used in this experiment.

[2] Requests for reprints should be sent to N. H. Kirby, Department of Psychology, University of Adelaide, Adelaide, South Australia 5001, Australia.

Apparatus.—The stimuli were presented by means of two lights of 16-mm. diameter set 6 mm. apart and mounted on a blackboard with a white vertical line between them. The board was situated 2.8 m. from *S*. Responses were made with two flat-topped telegraph keys placed on a board in front of *S*. The left index finger operated the left-hand key in response to the left light and the right index finger the right-hand key for the right light. A pressure of approximately 140 gm. was required to operate the keys. Reaction time was recorded in milliseconds from the onset of the stimulus to the depression of either of the keys. The interstimulus intervals were 2, 4, and 8 sec. Recording equipment was situated in a room adjacent to that in which the experiment was conducted. Stimuli were presented and responses were recorded by means of computer-programmed paper tapes.

Stimulus sequences.—Six different runs of 100 trials each were prepared from random-number tables with constraints ensuring that both stimulus lights were presented equally often and that at least two examples of each possible sequence of five stimuli occurred in each run.

Procedure.—Each *S* was assigned to one of the six possible orders of the three interstimulus intervals and attended on 3 consecutive days, completing the six different runs of stimuli at a different interstimulus interval on each day. Thus the 8-sec. interstimulus interval session lasted for approximately 1.5 hr., the 4 sec. for 1 hr., and the 2 sec. for .5 hr. Within each session, there was a rest period of approximately 2 min. between each of the six runs. The *S*s were informed of the interstimulus interval before each session. They were then told that both the stimuli would be presented randomly an equal number of times in each 100 trials and were instructed to respond as rapidly as possible while keeping their error rate down to less than 5%.

RESULTS

Initial inspection of the results indicated that those of the first one or two runs differed from those of later runs. Accordingly, the first two and the final four runs of each interstimulus interval condition were combined and the mean RTs calculated for each possible combination of stimuli up to and including the fifth order. The lowest number of responses upon which a fifth-order mean was based for the first two runs combined was 25, most being calculated from approximately 30 responses. Lower order mean RTs were, of course, based upon substantially larger numbers of responses. The means for the last four runs combined were based upon approximately twice as many responses as for the first two runs. These results are set out graphically in the form of tree diagrams in Fig. 1, where each node represents the mean RT for a particular combination of stimulus events. The single letter A represents the overall mean RT. AA and BA indicate the RTs for a stimulus preceded by the same and different stimuli, respectively, and similarly for the third-, fourth-, and fifth-order sequences. Errors amounted to approximately 2% of the total number of responses, and were included in the analyses as if they had been correct responses to stimuli which were other than those which had in fact occurred.

The results shown in Fig. 1 reveal both second-order and higher order sequential effects giving faster *alternations* than repetitions. The effects are of a similar pattern to those found by Remington (1969) except that in his experiment repetitions were faster than alternations. RTs decrease as the length of an alternation sequence increases, that is, from BA to ABA to BABA to ABABA. Similarly, for the upper branch of the tree diagrams it can be seen that a stimulus which does not continue the alternation sequence is responded to relatively slowly and that this slowness generally increases with the length of the preceding alternation sequence, that is, from AA to BAA to ABAA to BABAA.

A three-way analysis of variance was performed on the differences between AA and BA across the three interstimulus-interval conditions and over runs (Interstimulus Interval × Number of Runs × *S*s). There was no significant effect due to interstimulus interval but there was a significant effect due to number of runs, $F\ (1,\ 5) = 26.68,\ p < .005$. From the diagrams, this decrease from the first two to the last four runs does not appear to be very marked except in the 8-sec. condition. However, there was no significant interaction between interstimulus interval and the number of runs.

In order to test the significance of the higher order alternation effects, a five-way analysis of variance was performed using as the raw scores the differences between the RTs of alternation sequences of differ-

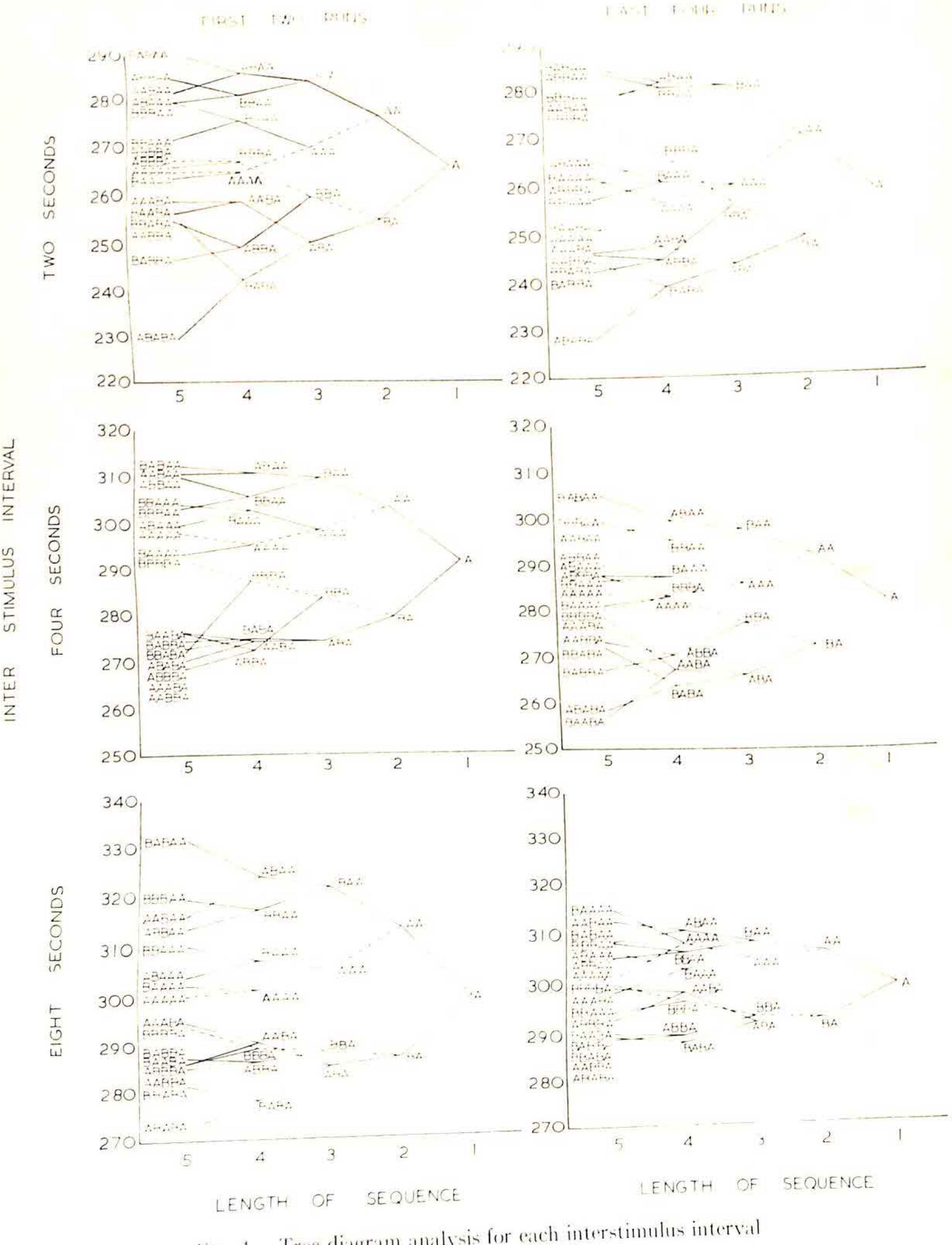


FIG. 1. Tree diagram analysis for each interstimulus interval and for the first two and last four runs.

ent lengths (i.e., ABA–BA, BABA–ABA, and ABABA–BABA) and the differences between the RTs of the discontinuations of alternation sequences of different lengths (i.e., BAA–AA, ABAA–BAA, and BABAA–ABAA). Other factors in the analysis were

*S*s, first two versus last four runs, interstimulus interval, and a comparison between the differences involving alternation sequences and those involving the discontinuations of alternation sequences. This last comparison was the only significant main effect, $F\ (1, 5) = 200.42$, $p < .005$. The only significant interaction was that due to the Significant Main Effect × Interstimulus Interval, $F\ (2, 10) = 9.802$, $p < .005$. These results indicate that higher order alternation effects do occur and that they change with increasing interstimulus interval. From the diagrams, this decrease in the higher order alternation effects with increasing interstimulus interval appears to be confined to the last four runs. However, the interaction corresponding to this was not found to be significant.

Sequences of *repetitions* and the discontinuation of such sequences are shown by the dotted lines in the diagrams. It can be seen that, generally, the RT decreases as the length of the repetition sequence increases and that the opposite occurs with increasing length of repetition sequence preceding a discontinuation.

In order to test the higher order repetition effects, the same five-way analysis of variance as that used for the higher order alternation effects was performed except that in this case the differences between repetition sequences of different lengths and the differences between the discontinuations of repetition sequences of different lengths were used as the raw scores (i.e., AAA–AA, AAAA–AAA, AAAAA–AAAA; and BBA–BA, BBBA–BBA, BBBBA–BBBA). The comparison between differences involving repetition sequences and those involving the discontinuations of repetition sequences was found to be significant, $F\ (1, 5) = 17.05$, $p < .025$. There were no other significant effects.

DISCUSSION

Williams' (1966) model for the alternation effect assumes a comparison system which matches each new input against the input on the previous trial. Comparisons resulting in a "match" decision are assumed to require more time than comparisons resulting in a "non-match" decision. It is clear from the present results, however, that this model is inadequate insofar as it assumes that only the immediately preceding stimulus is important. Remington (1969) dismissed Bertelson's 1963 model for the repetition effect for the same reason. A parsimonious explanation for both the present results and those of Remington for repetitions can be offered in terms of flexible guessing strategies as suggested by Keele (1969) and by Schvaneveldt and Chase (1969), perhaps involving some kind of preparation which can be increased according to the number of previously correct anticipations. Thus, the longer the preceding sequence of either repetitions or alternations, the shorter would be the RTs for stimuli which continued the sequence, and the longer would be the RTs for stimuli, which discontinued the sequence. There might perhaps be a limit after which *S* would begin to expect a change from the repetition or alternation sequence, but this was not apparent in the present results for the alternation effect, nor in Remington's for the repetition effect, for sequences of up to five in length.

The alternation effects in the present experiment could be explained by assuming the *S*s tended to begin with an initial bias in favor of more alternations than repetitions perhaps analogous to the gambler's fallacy phenomenon.

The fact that, overall, the RTs for repetition sequences decrease as the length of the sequence increases and that the opposite occurs for the discontinuations of such sequences could be explained by further assuming that *S*s on some occasions are prepared to change their strategy from preparation for alternations to preparation for repetitions after the first two or three of a run of the same stimulus, reverting back, however, to the alternation strategy after the occurrence of the next alternation.

The overall decrease from the first two to the last four runs in the second-order alternation effect might perhaps be due in part to *S* changing his strategy toward more preparation for repetitions, perhaps as a result of becoming more aware that there are equal numbers of repetitions and alternations. The slight increase in the higher order *repetition* effects across runs in the 2-sec. condition would be consistent with this. In the 8-sec. condition, the decrease in the second-order alternation effect and the apparent decrease (albeit nonsignificant) in the higher order alternation effects from the first two to the last four runs might have been due to boredom, causing *S*s to make less thorough preparation for their reactions. The *S*s did report that they found the 1.5 hr. in the 8-sec. condition extremely tiring.

A decrease in *S*'s preparation due to boredom might also account in part for the overall decrease in the higher order alternation effects with increasing interstimulus interval. However, this result might also be expected on the above explanation insofar as stimuli further back in the sequence would be likely to have less effect the further removed in time they are from the present stimulus, since *S*'s expectation and preparation are presumably based on memory traces of the preceding stimuli which might be assumed to decay with time. On this explanation, one might also expect that the higher order repetition effects would decrease with increasing interstimulus interval. While there is some suggestion from the diagrams that this is so, the effect was not found to be significant, perhaps because the repetition effects are not as marked in any of the conditions as the alternation effects.

Remington (1969) obtained a repetition effect with a 4-sec. interstimulus interval while the present experiment found alternation effects for 2-, 4-, and 8-sec. interstimulus intervals. Since there appears to be no obvious difference in the procedures used that might account for this difference, it suggests that the repetition and alternation effects are not definitely confined to a particular range of interstimulus intervals. If so, this fact would appear to lend further support to an explanation of alternation and repetition effects in terms of flexible strategies which *S* may vary according to circumstances. The factors influencing such strategies are not at present clear and require further investigation.

REFERENCES

BERTELSON, P. Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, 1961, **13**, 90–102.

BERTELSON, P. S–R relationships and reaction times to new versus repeated signals in a serial task. *Journal of Experimental Psychology*, 1963, **65**, 478–484.

HALE, D. J. Sequential effects in a two-choice serial reaction task. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 133–141.

KEELE, S. W. Repetition effect: A memory-dependent process. *Journal of Experimental Psychology*, 1969, **80**, 243–248.

REMINGTON, R. J. Analysis of sequential effects in choice reaction times. *Journal of Experimental Psychology*, 1969, **82**, 250–257.

SCHVANEVELDT, R. W., & CHASE, W. G. Sequential effects in choice reaction time. *Journal of Experimental Psychology*, 1969, **80**, 1–8.

WILLIAMS, J. A. Sequential effects in disjunctive reaction time: Implications for decision models. *Journal of Experimental Psychology*, 1966, **71**, 665–672.

(Received December 28, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 37–42

FAMILIAR AND RELATIVE SIZE CUES AND SURFACE TEXTURE AS DETERMINANTS OF RELATIVE DISTANCE JUDGMENTS[1]

COLIN V. NEWMAN[2]

University of Birmingham, England

A series of experiments showed that familiar size and relative size cues to distance, either separately or in combination, influenced relative distance judgments in which *S* attempted to position one object midway between two other objects. These were placed at different physical distances from *S*, on a surface that was either plain white or clearly textured. The *S* viewed the objects under reduced, monocular conditions. When the information to distance in the surface texture was held constant, relative size cues influenced relative distance judgments. When, however, familiar size and relative size cues were held constant, relative distance judgments were influenced by variations in surface texture. Results indicated that J. J. Gibson may have placed undue stress on the importance of the texture variable in formulating his psychophysical hypothesis.

Hochberg and McAlister (1955) argued that the familiar size cue to relative distance can be subsumed under the relative size cue to relative distance. More recently, both Epstein (1967) and Gogel (1964) concluded that in some experiments familiar size has been shown to act as a cue to relative distance independently of relative size. This has been confirmed by Gogel and Mertens (1968) and Ono (1969). Past experiments have investigated the influence of familiar and relative size cues on the perceived relative distance of, typically, just two objects viewed in environments devoid of any other distance information. To test the functional validity of these cues under less restricted conditions, the present series of experiments examined whether familiar and relative size cues influenced the judged distance of one object with respect to two others, when all three objects were physically at different distances from *S* and were placed on a surface that was either plain white or strongly textured.

Gibson (1950) accepted that *S*'s explicit judgments in an experiment may be influenced by subjective assumptions about the familiar or relative sizes of objects when *E* has deliberately removed all normally available sources of stimulation, such as texture density gradients. When such gradients of stimulation are present, Gibson argues that *S*s do not make assumptions about objects in the environment since the necessary information about their distances and sizes is directly available in the pattern of stimulation on the retina. Two experiments in the present series tested Gibson's view by examining whether judgments of relative distance, made over a surface with a well-defined texture, were influenced by texture alone, or by the presence of conflicting familiar and relative size cues to distance.

EXPERIMENT I

Experiment I tested the hypothesis that in combination, familiar and relative size cues would influence relative distance judgments on a textureless surface.

Method

Apparatus.—Figure 1 shows a schematic diagram of the apparatus.

The apparatus was also used in all subsequent experiments and has been fully described by Newman (1971). It featured a viewing box, 152.4 cm. wide × 76.2 cm. long, into which *S* looked through a monocular aperture at the center of the front end

---

[1] The research was conducted at the University of Leicester in partial fulfillment of the requirements for a PhD degree. The author held a Social Science Research Council Studentship. Acknowledgments are due to S. G. M. Lee for his advice on experimental design.

[2] Requests for reprints should be sent to Colin V. Newman, Department of Psychology, University of Birmingham, Birmingham B15 2TT, England.

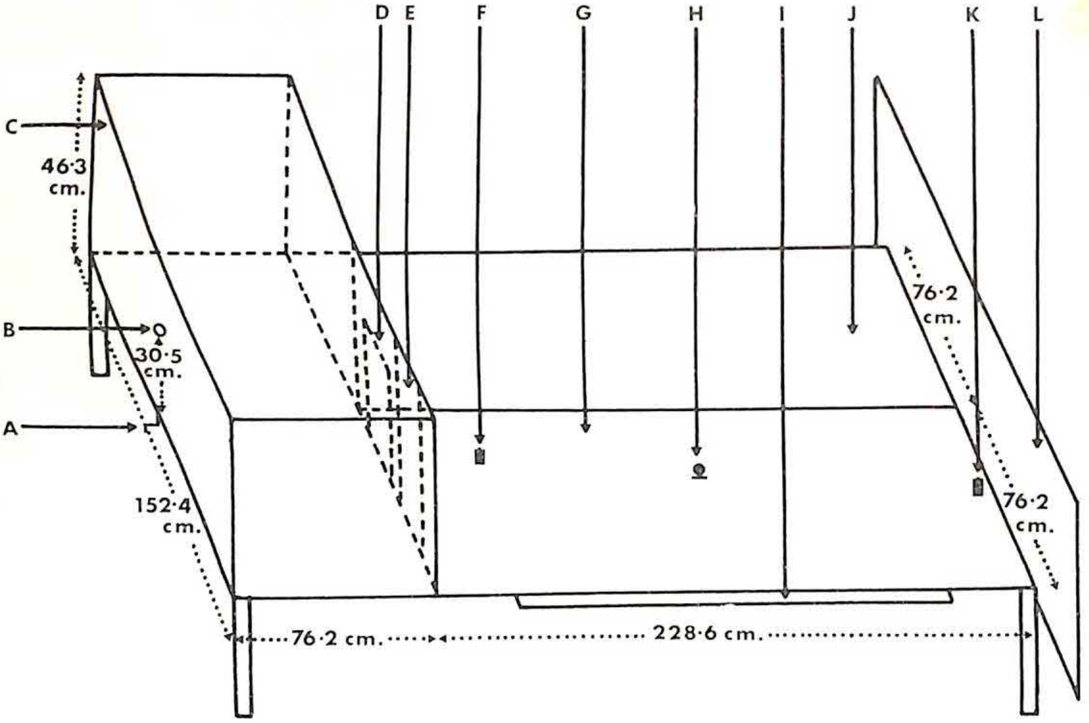


FIG. 1. Schematic diagram of apparatus: A = winding handle; B = monocular viewing aperture; C = viewing box; D = reduction screen with aperture; E = sliding door; F = near marker; G = plain white surface; H = distance bisection marker; I = measuring scale; J = alternative test surface, occluded throughout this series of experiments; K = far marker; L = white screen.

of the box, 30.5 cm. above its floor. When the door in a reduction screen facing *S* was opened, he looked along a plain white, horizontal surface 228.6 cm. long × 76.2 cm. wide, slightly to the right of *S*'s direct line of sight. The full width of the surface could not be seen by *S* as its edges were occluded by the reduction screen. Vision beyond the end of the surface was restricted by a white screen, mounted vertically. Two markers were placed on the surface. These defined a distance which *S* was to bisect. The "far marker" was positioned at the horizon end of the surface and the "near marker" 30.5 cm. from the reduction screen aperture. The near marker was a standard-sized, familiar flashlight battery (5.8 cm. tall × 3.3 cm. diam.), and the far marker was either a second standard-sized flashlight battery or an off-sized, scaled miniature of this flashlight battery (4.4 cm. tall × 2.5 cm. diam.). A distance bisection marker, consisting of a standard-sized table tennis ball (3.8 cm. in diam.) riveted to a magnetic disc (2.8 cm. in diam.), was placed on the surface above a powerful button magnet, mounted on tracks and held immediately under the test surface. By winding a handle, *S* could vary the position of the bisection marker along the direct line of sight between near and far markers. The position of the bisection marker could be read from a scale mounted beside the test surface and visible only to *E*.

Procedure.—To remind *Ss* of the standard sizes of the specific, familiar objects to be viewed subsequently, all *Ss* were given the relevant familiar objects to hold before the experiment. Twenty adult *Ss* of both sexes, with normal or corrected-to-normal vision, then made distance bisection judgments. For half of the *Ss* the standard-sized flashlight battery was the far marker on the first four trials and the off-sized flashlight battery was the far marker on the final four trials. The remaining *Ss* made the same judgments, but in reverse order. On all eight trials, *Ss* were instructed to move the distance bisection marker to the physical midpoint between the far and near markers. The psychophysical method of adjustment was used. In each condition, *S*'s point of subjective equality was calculated as the mean of his four judgments of the midpoint. The crossover (Latin square) design was employed to balance the order in which bisection judgments were made under the two experimental conditions.

Rationale and Prediction

Provided that *Ss* assume all objects, including the off-sized flashlight battery, have the standard sizes of familiar objects, it is predicted that the reduced size of the far object would increase its apparent distance and hence the apparent halfway distance as well. If familiar size and relative size together act as cues to relative distance, then *S* will move the bisection marker until it is positioned objectively too near to the far off-sized flashlight battery. When the far marker is changed to a standardized flashlight battery, *S* should place the table tennis ball nearer to the objective midpoint.

Results and Discussion

Informal questioning of each *S* at the end of the experiment indicated that no *S* had detected the distorted size of the far flashlight battery

Judgments of the midpoint were expressed as deviations from the objective midpoint; positive deviations were recorded when the bisection marker was placed objectively too near to the far marker, negative deviations in the reverse situation. Mean errors in distance bisection for all *S*s are shown in Table 1.

The hypothesis was confirmed: Positive errors in distance bisection were significantly greater with the off-sized rather than standard-sized flashlight battery as the far marker, $F(1, 18) = 5.62$, $p < .05$. There was no significant difference between groups of *S*s (A vs. B) or between judgments made first and judgments made second.

This experiment did not test the separate influences of familiar size and relative size on relative distance judgments. This was done in Exp. II and III.

TABLE 1

MEAN ERRORS IN DISTANCE BISECTION ON A TEXTURELESS SURFACE WITH FAMILIAR AND RELATIVE SIZE CUES PRESENT

Group of *S*s	$\bar{X}$ and SD, first four judgments	$\bar{X}$ and SD, second four judgments
A	16.0 cm.^{a} ±11.13 cm.^{a}	22.5 cm.^{b} ±17.90 cm.^{b}
B	22.2 cm.^{b} ±11.65 cm.^{b}	16.1 cm.^{a} ±11.35 cm.^{a}

Note.—Group A ($n = 10$); Group B ($n = 10$).
[a] Standard-sized, far flashlight battery.
[b] Off-sized, far flashlight battery.

TABLE 2

MEAN ERRORS IN DISTANCE BISECTION ON A TEXTURELESS SURFACE WITH FAMILIAR SIZE CUES PRESENT

Group of *S*s	$\bar{X}$ and SD, first four judgments	$\bar{X}$ and SD, second four judgments
A	14.9 cm.^{a} ±12.17 cm.^{a}	22.0 cm.^{b} ±12.91 cm.^{b}
B	17.8 cm.^{b} ±17.72 cm.^{b}	17.6 cm.^{a} ±16.51 cm.^{a}

Note.—Group A ($n = 16$); Group B ($n = 16$).
[a] Standard-sized, far flashlight battery.
[b] Off-sized, far flashlight battery.

EXPERIMENT II

This experiment tested the hypothesis that familiar size, independently of relative size, would influence relative distance judgments on a textureless surface.

Method

The experiment was an exact replication of Exp. I except for the following changes: The near flashlight battery was replaced by a standard-sized matchbox of a common "Swan Vesta" brand (7.9 cm. wide × 4.9 cm. tall). Sixteen, instead of 10 *S*s, were included in each experimental group, none having served in the previous experiment.

Rationale and Prediction

When *S*s observe unfamiliar objects of different shapes there is no basis for them to assume that "similar things are identical [Ittelson, 1960, p. 70]." It is thus normally assumed that relative size operates as a cue to relative distance only when comparison objects have the same shape (Gogel, 1964; Hochberg & McAlister, 1955). Hence, in this experiment, when differently shaped familiar objects were viewed, familiar size cues were isolated from relative size cues; the latter were assumed to be inoperative, as all objects had different shapes. On the basis of this assumption the experimental hypothesis was tested.

Provided all objects in *S*s' field of view, including the off-sized far flashlight battery, are assumed by *S*s to have the standard sizes of familiar objects, the following hypothesis is advanced on the same basis as in Exp. I. If familiar size serves as a cue to relative distance, the table tennis ball will be placed objectively nearer to the far marker when this is an off-sized flashlight battery than when it is a standard-sized battery.

Results and Discussion

Informal questioning at the end of the experiment again indicated that no *S* had detected any size distortion of the far flashlight battery. Mean errors in distance bisection for all *S*s are shown in Table 2.

The hypothesis was confirmed: Positive errors in distance bisection were significantly greater with the off-sized rather than the standard-sized flashlight battery as the far marker, F $(1, 30) = 9.24$, $p < .01$. There was no significant difference between groups of *S*s (A vs. B), but the difference between judgments made first and judgments made second was significant, F $(1, 30) = 10.26$, $p < .01$. This may reflect a repeated judgment effect (Newman, 1968; Wohlwill, 1964). Repeated bisection judgments result in some *S*s placing the bisection marker increasingly nearer the far marker on successive trials.

The finding of Gogel and Mertens (1968) was confirmed by the result which showed that familiar size served as a cue to relative distance between the differently shaped familiar objects.

EXPERIMENT III

This experiment tested the hypothesis that relative size, independently of familiar size, would influence relative distance judgments despite the presence of clear distance information in the texture on the surface over which judgments were made.

Method

The experiment was an exact replication of Exp. I, again testing 20 *S*s, except that the following changes in the apparatus were made. The plain white surface was replaced by a regular textured surface having the same physical dimensions, but painted crosswise to *S*'s line of sight with 63 alternate black and white stripes, each 3.63 cm. wide. The standard-sized flashlight batteries were replaced by blue cardboard rectangles, held vertically from behind, all being 5.8 cm. tall × 3.3 cm. wide. The off-sized far flashlight battery was replaced by a smaller blue cardboard rectangle, 4.4 cm. tall × 2.5 cm. wide, being a scaled miniature of the larger rectangles, but having the same outline shape. To maintain consistent conditions between experiments, the table tennis ball was retained as the distance bisection marker.

Rationale and Prediction

Provided all rectangles, including the small rectangle, are assumed equal in physical size, then, if relative size serves as a cue to relative distance, the bisection marker will be placed objectively nearer to the far marker when this is the 4.4-cm.-tall rectangle than when it is the 5.8-cm.-tall rectangle. The textured surface was introduced to test whether relative size is a functional cue to relative distance when surface texture provides clear information about the objective midpoint. Gibson (1959) argued that *S*'s judgment of the length of one stretch of ground relative to the other depends on the number of transitions or textural units in one stretch relative to the number of similar units in the other. Therefore, when making bisection judgments on a textured surface, *S*s should attempt to equate the number of textural units beyond the midpoint with the number of similar units in front of it. Hence, on this surface, under all conditions, the texture should enhance objectively correct judgments of the midpoint.

Results and Discussion

Informal questioning at the end of the experiment indicated that all *S*s had assumed that the 5.8-cm.-tall rectangles and

TABLE 3

MEAN ERRORS IN DISTANCE BISECTION ON A REGULAR TEXTURED SURFACE WITH RELATIVE SIZE CUES PRESENT

Group of *S*s	$\bar{X}$ and *SD*, first four judgments	$\bar{X}$ and *SD*, second four judgments
A	13.5 cm.[a] ±14.05 cm.[a]	18.5 cm.[b] ±17.27 cm.[b]
B	20.9 cm.[b] ±10.80 cm.[b]	19.2 cm.[a] ±11.82 cm.[a]

Note.—Group A ($n = 10$), Group B ($n = 10$).
[a] 5.8-cm. rectangle as far marker.
[b] 4.4-cm. rectangle as far marker.

the 4.4-cm.-tall rectangle were equal in both physical size and shape. Mean errors in distance bisection for all *S*s are shown in Table 3.

Positive errors in distance bisection were significantly greater with the 4.4-cm. rectangle rather than the 5.8-cm. rectangle as the far marker, F (1, 18) = 5.60, $p < .05$. There was no significant difference between judgments made first and judgments made second, but there was a significant difference between groups of *S*s (A vs. B), F (1, 18) = 8.34, $p < .05$. This may reflect an interaction between *S*s and the repeated judgment effect discussed earlier.

The results showed that relative size served as a cue to distance, despite the presence of the information to distance in the surface texture. It cannot be concluded that texture had no influence on distance judgments. Experiment IV directly tested the influence of texture on relative distance judgments when constant familiar size and relative size cues were also present.

EXPERIMENT IV

Newman (1971) has shown that a progressive variation in the packing density of the texture elements on a surface along its length influenced *S*'s judgments of the midpoint. This experiment tested the hypothesis that such physical distortions in the texture on a surface would influence judgments of the midpoint, despite the presence of familiar and relative size cues to distance.

Method

The regular textured surface from Exp. III and a further decreasing textured surface were used on the same apparatus. The decreasing textured surface was cut from hardboard to the same physical dimensions as the regular textured surface, but the 63 alternating black and white stripes were painted so that the stripe nearest to *S*'s viewing position was 5.08 cm. wide and each successive stripe was 1/90 narrower than the previous one, making the final horizon stripe 2.50 cm. wide. For all judgments, standard-sized flashlight batteries were used as the far and near markers and the table tennis ball as the distance bisection marker. Sixteen naive adult *S*s made distance bisection judgments exactly as in Exp. I, but first on the decreasing textured surface and then on the regular textured surface. The crossover design was abandoned since these *S*s were also to serve in a different experiment in which common prior experience was essential. The sequence in which judgments were made was arranged so that any repeated judgment effects would militate against confirmation of the experimental hypothesis.

Rationale and Prediction

If *S*s attempt to equate the number of textural units in front of the midpoint with the number of similar units behind it, the mean distance bisection score on the decreasing textured surface is greater than on the regular textured surface (Newman, 1971). If, however, familiar and relative size cues to relative distance overrule the influence of texture, a difference in these means is not found.

Results and Discussion

The mean error in distance bisection on the decreasing textured surface was 19.5 cm. ± 14.98 cm. and on the regular textured surface was 14.7 cm. ± 15.53 cm. These means were significantly different, for related samples, t (30) = 4.28, and r (15) = .96, both ps $< .01$.

DISCUSSION AND SUMMARY

Familiar size and relative size cues to distance, either separately or in combination, influenced relative distance judgments of the position of one object between two others when all objects were at different physical distances from *S*. When the information as to distance in the texture on the surface over which judgments were made was held constant, relative size cues influenced relative distance judgments. When, however, familiar size and relative size cues were held constant, relative distance judgments were influenced by variations in surface texture.

The finding that relative size cues influenced bisection judgments when surface texture also provided clear information about the objective position of the midpoint does not seriously challenge Gibson's (1950) psychophysical hypothesis. He accepts that a single, static, monocular texture density gradient, like that available to *S*s in these experiments, is an ambiguous stimulus corresponding to a family of tridimensional arrangements of surface slant and spacing of the texture elements (Gibson, Purdy, & Lawrence, 1955). Until this ambiguity is removed by the presence of other gradients of stimulation, *S*s can, at best, only assume that the physical sizes of successive texture elements are equal. Such an assumption, made by *S*s in one particular experimental situation, might be abandoned in favor of alternative assumptions concerning the physical sizes of objects on the test surface when situational variables in the experiment are altered, and distance information in the surface texture remains ambiguous. Gibson's theory would be challenged if the familiar and relative size cues, found operative in this experiment, are still found to influence relative distance judgments when these are made in unrestricted, natural environments in which *S*s are permitted binocular vision. In these conditions, the distance information in texture density gradients is unambiguous.

The present results confirm that a single, static, monocular texture density gradient cannot unequivocally determine relative distance judgments. Other sources of distance information were found to influence distance judgments even when a well-defined texture density gradient was present. This suggests that texture density gradients may have been afforded a place of undue importance in initial formulations of Gibson's (1950) theory.

REFERENCES

EPSTEIN, W. *Varieties of perceptual learning.* New York: McGraw-Hill, 1967.

GIBSON, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.

GIBSON, J. J. Perception as a function of stimulation. In S. Koch (Ed.), *Psychology, a study of a science.* Vol. 1. New York: McGraw-Hill, 1959.

GIBSON, J. J., PURDY, J., & LAWRENCE, L. A method of controlling stimulation for the study of space perception: The optical tunnel. *Journal of Experimental Psychology*, 1955, **50**, 1–14.

GOGEL, W. C. Size cue to visually perceived distance. *Psychological Bulletin*, 1964, **62**, 217–235.

GOGEL, W. C., & MERTENS, H. W. Perceived depth between familiar objects. *Journal of Experimental Psychology*, 1968, **77**, 206–211.

HOCHBERG, J. E., & MCALISTER, E. Relative size vs. familiar size in the perception of represented depth. *American Journal of Psychology*, 1955, **68**, 294–296.

ITTELSON, W. H. *Visual space perception.* New York: Springer, 1960.

NEWMAN, C. V. The influence of both visual texture density gradients and object familiarity on relative size and distance judgments. Unpublished doctoral dissertation, University of Leicester, England, 1968.

NEWMAN, C. V. The influence of visual texture density gradients on relative distance judgments. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 225–233.

ONO, H. Apparent distance as a function of familiar size. *Journal of Experimental Psychology*, 1969, **79**, 109–115.

WOHLWILL, J. F. Changes in distance judgments as a function of corrected and noncorrected practice. *Perceptual and Motor Skills*, 1964, **19**, 403–413.

(Received January 3, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 43–48

EFFECTS OF PRIOR UNCERTAINTY ON INCIDENTAL FREE RECALL[1]

D. E. BERLYNE[2] AND LORRAINE F. NORMORE

University of Toronto, Ontario, Canada

Subjective uncertainty, induced through blurred pictures, was found to facilitate both immediate and delayed incidental recall, but not intentional recall, of objects portrayed in subsequent clear pictures. This finding corroborated those of previous experiments, in which the induction of uncertainty through prior guessing facilitated incidental paired-associate learning of words. Further experiments showed the effect to be absent when blurred pictures were dissociated from uncertainty or when clear pictures did not relieve uncertainty. Finally, the effect occurred, whether or not *S* knew, on seeing a blurred picture, that the corresponding clear picture would follow. Explanations in terms of reinforcement through reduction of epistemic curiosity and in terms of attention were discussed.

In a series of experiments on paired-associate learning by Berlyne, Carey, Lazare, Parlow, and Tiberius (1968), the stimulus terms were Turkish words and the response terms were corresponding English words. Incidental learning was improved, but intentional learning impaired, by presenting the Turkish word and requiring *S* to guess its meaning just before the appearance of the two words together. Control experiments showed that the facilitation of incidental recall depends on the induction of subjective uncertainty during the guessing phase and on the receipt of information eliminating the uncertainty during the learning phase. The findings were therefore compatible with the view that subjective uncertainty gives rise to an aversive state (epistemic curiosity), whose relief through exposure to relevant information is a reinforcing condition (Berlyne, 1954, 1965). An alternative interpretation is that uncertainty causes *S* to attend more intensely to a response term bearing relevant information, so that its retention is favored.

The experiments reported in the present article pursued this line of research with two major changes in technique. First, incidental free recall of pictured objects was studied rather than paired-associate recall of words. Second, subjective uncertainty was induced by exposing a blurred picture immediately before a clear version of the same picture. Experiments by Nicki (1968, 1970) indicated that subjective uncertainty is an inverted U-shaped function of degree of blur. He also showed that the presentation of a clear picture after its blurred counterpart reinforces a manual response more effectively than the sight of an unrelated clear picture. Berlyne and Borsa (1968) demonstrated that blurred pictures induce longer EEG desynchronization than clear pictures, but only when *S*s are uncertain about the objects portrayed in the blurred pictures. The stimulus material used in the present experiments was selected from the colored slides used by Berlyne and Borsa and the black-and-white slides used by Nicki (1968, 1970). The degree of blur was identical with, or close to, the degree that had produced maximum uncertainty and maximum reinforcing effect in Nicki's experiments.

EXPERIMENTS I, II, AND III

The main purpose of Exp. I was to compare immediate incidental recall after 5 sec. of exposure to a clear picture of an

---

[1] This investigation was supported by Research Grants APA-73 and A-73 from the National Research Council of Canada. The data were reported by L. F. Normore in a thesis submitted in accordance with the requirements for the master's degree of the University of Toronto. The experiments were suggested partly by some data that Donna M. Borsa collected.

[2] Requests for reprints should be sent to D. E. Berlyne, Department of Psychology, University of Toronto, Toronto 5, Ontario, Canada.

TABLE 1

Mean Number of Items Recalled, Experiments I–V

Cond.	Exp.				
	I (imm., inc.)	II (del., inc.)	III (imm., int.)	IV (imm., inc.)	V (imm., inc.)
Single-Clear	2.5	1.6	4.6		
Double-Clear	3.4	2.6	5.2	2.8	4.0
Blurred-Clear	4.0	3.2	5.2	3.8	4.2
Clear-Blurred				2.7	
Unrelated Blurred-Clear					3.0

Note.—Imm. = immediate recall, del. = delayed recall, inc. = incidental learning, and int. = intentional learning. Total no. of items recalled = 8.

object with recall after 5 sec. of exposure to a clear picture preceded by 5 sec. of exposure to a blurred picture of the same object. However, any superiority of the latter condition could be attributed to greater duration of total exposure. Consequently, a third condition was introduced, consisting of the appearance of a clear picture for 10 sec.

Experiments II and III were identical with Exp. I, except for the following differences. In Exp. II, there was a 24-hr. delay between the presentation of the items and the recall tests. The *S*s were told at the beginning of the experiment that two recordings of galvanic skin response (GSR), 24 hr. apart, would be needed to investigate shifts over time. In Exp. III, immediate intentional recall was studied.

Method

Subjects.—The *S*s of Exp. I and II consisted of undergraduates taking elementary and intermediate psychology courses. There were 21 males and 21 females in Exp. I and another 21 males and 21 females in Exp. II. Eleven male and 13 female summer-session undergraduates taking psychology courses participated in Exp. III.

Data from five additional *S*s in Exp. I were discarded because, when questioned after the experiment, they said that they had tried to remember the objects while looking at the slides. In Exp. II, all of the *S*s who reported any effort to remember the objects during the 24-hr. interval between the training and test sessions were discarded. There were 49 of them.

Stimulus material.—Twenty-four pairs of colored slides were used, each pair consisting of a clear picture of a common object on a plain background and a blurred version of the same picture.

Experimental design.—The 24 items were presented to all *S*s in the same order. One-third of the items appeared in each of the following conditions: (*a*) Single-Clear (the screen blank for 5 sec., followed immediately by a clear slide for 5 sec.), (*b*) Double-Clear (a clear slide for 5 sec., followed immediately by the same clear slide for 5 sec.), and (*c*) Blurred-Clear (a blurred slide projected for 5 sec., followed by a corresponding clear slide for 5 sec.).

Of each consecutive set of three items, one item was presented in each of the three conditions with the order randomized. The *S*s were divided into three groups of 14, such that each item was presented in the three conditions to different groups.

Procedure.—The *S*s participated individually. They were told that the experiment was designed to investigate psychophysiological reactions to pictures. A GSR electrode connected with a lead going to a patch panel in the wall was attached to *S*'s dominant hand. It was explained to *S* that he would sometimes see a clear picture alone, sometimes a clear picture preceded by a blurred version of the same picture, and sometimes a clear picture of the same object twice. The *S*s of Exp. III were given the same information, but they were also told, before the slides were presented, that they should try to remember as many as possible of the objects depicted and that a recall test would be given immediately afterwards.

The slides were projected by a Kodak Carousel Model 550 projector, located just behind *S*'s right shoulder, onto a screen 2.5 m. away. The size of the image on the screen was 66 × 44 cm.

After the slides had been presented, the electrode was removed, and *S* was asked to write down the names of as many objects as he could recall from the slides he had just seen. He was subsequently asked whether he had tried to guess the identity of the objects depicted in the blurred slides, and whether he had tried to remember the objects depicted in the clear slides.

Results

The mean numbers of items seen in each of the three conditions that were recalled by the *S*s of the three experiments are displayed in Table 1.

Experiment I (immediate recall, incidental learning).—The means for the three conditions were significantly different, $F(2, 82) = 12.86$, $p < .001$. A Duncan new multiple-range test showed the Blurred-Clear mean to be significantly higher ($p < .05$) than the Double-Clear mean, which in its turn was higher ($p < .01$) than the Single-Clear mean.

The findings thus favor the same conclusion as those of Berlyne et al. (1968):

The induction of uncertainty before the presentation of the information to be retained produces a greater improvement in incidental recall than doubling the length of time for which the information is presented.

Experiment II (delayed recall, incidental learning).—The means were significantly different, $F\ (2, 82) = 18.53$, $p < .001$. Duncan's test showed the Blurred-Clear mean to be higher than the Double-Clear mean ($p < .05$), and the Double-Clear mean to be higher ($p < .001$) than the Single-Clear mean.

Experiment III (immediate recall, intentional learning).—The Blurred-Clear and Double-Clear means were close together, and the Single-Clear mean was somewhat lower. Analysis of variance revealed, however, no significant difference. We have therefore no evidence that the uncertainty-inducing treatment favors intentional recall.

EXPERIMENT IV

In order to verify that the superiority of incidental recall after the Blurred-Clear condition was due to the induction of uncertainty, it was necessary to see what would happen if *S*s were given equal exposure to the blurred and the clear versions of a picture in circumstances where the appearance of the blurred slide would not induce uncertainty. The blurred slide was robbed of its association with uncertainty by a method that Berlyne and Borsa (1968), as well as Nicki (1970), used to verify that effects they were studying depended on uncertainty and not on blur as such: The blurred slide was presented after the corresponding clear slide.

Method

Subjects.—There were 10 male and 10 female summer-session undergraduates. Two additional *S*s were discarded because they reported having made an effort to remember the stimulus material.

Experimental design and procedure.—Twenty-eight colored slides picturing common objects on a plain background were used. The 28 items appeared in the same sequence for all *S*s, and the three treatments were rotated among items in the following manner. For each consecutive pair of 4 items, 2 were presented in the Double-Clear condition, 1 in the Blurred-Clear condition, and 1 in the Clear-Blurred condition (i.e., 5 sec. clear followed by 5 sec. blurred). Thus, each item was used in the Double-Clear condition for 10 *S*s and in each of the other two conditions for 5 *S*s. These two restrictions were ensured with the help of four randomly selected 4 × 4 Latin squares.

The *S*s were told beforehand that clear pictures would sometimes be preceded, and sometimes followed, by blurred versions of the same picture. In order to make clear which blurred and clear slides were paired, the screen was left blank for 5 sec. between consecutive items. Once again, the GSR electrodes were applied, and *S*s were given no reason to expect a recall test.

Results

The means are displayed in Table 1, with the mean for the Double-Clear condition (which occurred twice as often as the other conditions) halved to make it comparable with the others. Analysis of variance showed the three means to be significantly different, $F\ (1, 32) = 3.36$, $p < .05$. Whereas in the previous experiments there were two control conditions, Single-Clear and Double-Clear, the Double-Clear treatment served as a sole control condition in this experiment. Consequently, Dunnett's (1955) test for comparing several treatments with a control treatment was applied. It showed that the Blurred-Clear mean was significantly higher than the Double-Clear mean at the .05 level. The Clear-Blurred mean was not significantly different from the Double-Clear mean and, in fact, slightly lower. Consequently, there is no indication that exposure to both blurred and clear versions of a picture facilitates incidental recall when the blurred version follows the clear version and *S* consequently knows what object it represents.

EXPERIMENT V

After Exp. IV had confirmed the importance of the uncertainty induced by the appearance of a blurred slide before a clear slide, Exp. V was carried out to investigate the necessity of the subsequent elimination of uncertainty by the appearance of a corresponding clear slide. In this experiment, the Double-Clear and Blurred-Clear conditions of previous experiments were combined with a new condition, Unrelated (Unrel.) Blurred-Clear, namely the presentation of a blurred slide followed by an unrelated clear slide.

Method

Subjects.—Data from 13 male and 17 female summer-session undergraduates were analyzed. Data from 10 additional *S*s were discarded because they reported efforts to learn the stimulus material.

Experimental design and procedure.—Twenty-eight different colored pictures of common objects on plain backgrounds served as stimulus material. Both blurred and clear versions of 21 of them were used, but only the blurred version of the remaining 7.

Each *S* had seven items in each of the three conditions. Two of them were Blurred-Clear and Double-Clear as in previous experiments. The third was Unrel. Blurred-Clear: A blurred slide was projected for 5 sec., followed by the clear version of a quite different slide for 5 sec.

In contrast with previous experiments, the three conditions were not interspersed but presented in successive blocks of seven items. This was necessary so that *S* could be informed about the conditions before each block. Without being told beforehand, *S* could not have known readily which blurred slides depicted the same objects as the clear slides that followed them and which did not. The 21 clear slides appeared in the same order for all *S*s. The three blocks corresponding to the treatments appeared in all six possible orders with equal numbers of *S*s.

Results

An analysis of variance showed significant main effects of treatments, F (2, 56) = 7.73, $p < .01$, and phases, F (2, 56) = 31.84, $p < .001$. The latter reflects the fact that the items occupying Positions 8–14 happened to be harder to remember than those coming earlier or later.

The means of the three conditions are shown in Table 1. According to Duncan's test, the difference between the Blurred-Clear and Double-Clear conditions, which had been significant in previous experiments, failed to reach significance. Both the Double-Clear and Blurred-Clear means in this experiment were comparable to the Blurred-Clear means obtained in Exp. I and IV. The grouping together of items belonging to the same condition that characterized the design of this experiment may conceivably have facilitated the retention of Double-Clear items. The Unrel. Blurred-Clear mean was, however, significantly lower than each of the other two means at the .001 level. So, we have strong evidence that the consecutive presentation of a blurred and a clear slide does not facilitate incidental recall unless the clear slide depicts the same object as the blurred slide and thus furnishes information capable of relieving the uncertainty induced by the latter.

EXPERIMENT VI

In all the experiments reported so far, whenever the Blurred-Clear condition occurred, *S*s knew that the blurred slide would be followed by a clear picture of the same object. The facilitation of incidental recall could therefore have been due to utilization of this forewarning. For example, the sight of a blurred slide might mobilize *S*'s attention, making him more receptive than he would otherwise be of the information contained in the clear slide. To test for this possibility, the final experiment used two groups of *S*s. For both groups, all items consisted of a blurred slide followed by a clear slide, but, in half of them, blurred and clear slides depicted the same object and, in the other half, they did not. The *S*s were informed which of these two conditions obtained, but, for one group, this information accompanied the onset of the blurred slide, whereas the other group did not receive it until the clear slide appeared.

Method

Subjects.—Data from 20 male and 20 female undergraduates, taking psychology courses during the academic year, were analyzed. Seventeen additional *S*s were discarded, 13 of them because they reported efforts to remember and 4 because they noticed that corresponding blurred and clear slides appeared in different Unrel. Blurred-Clear items.

Stimulus material.—Whereas colored slides had been used in the other experiments, 18 black-and-white pairs of slides were used in this one. The nine Unrel. Blurred-Clear items were formed by pairing each clear slide with the blurred version of another slide in the same set of nine. This ensured that both versions of every slide had one exposure in the Unrel. Blurred-Clear, as well as in the Blurred-Clear condition.

Experimental design and procedure.—The *S*s were divided into two groups known respectively as the

Before group and the After group. The same 24 items were used as in Exp. I. Half of the items were presented in the Blurred-Clear condition and half in the Unrel. Blurred-Clear condition, the two being interspersed in a randomized order. What was a Blurred-Clear item for one-half of each group of *S*s was an Unrel. Blurred-Clear item for the other half, and vice versa. It was explained to *S* beforehand that he would sometimes see a blurred picture followed by a clear version of the same picture and sometimes a blurred picture followed by a clear version of a quite different picture. A letter S (for "same") would denote the former condition and a letter D (for "different") the latter. The Before group were told that these letters would appear in the middle of the base of each blurred slide, and the After group in the same position on each clear slide. The blurred and clear slides were projected for 5 sec. each with a 5-sec. interitem interval.

TABLE 2

EXPERIMENT VI
MEAN NUMBER OF ITEMS RECALLED

Group	Cond.		
	Unrelated blurred-Clear	Blurred-Clear	Both
Clue on blurred slide	4.0	4.9	4.4
Clue on clear slide	3.3	4.5	3.9
Both	3.6	4.7	

Note.—Total no. of items recalled = 9.

Results

The means for the four conditions are displayed in Table 2. Analysis of variance revealed significantly better recall in the Blurred-Clear (S) than in the Unrel. Blurred-Clear (D) condition, F (1, 36) = 14.32, $p < .001$. There was no significant difference between the Before and After groups, F (1, 36) = 1.77, and no significant interaction between the Before/After and Blurred-Clear/Unrel. Blurred-Clear treatments, $F < 1$.

We therefore have further confirmation of the principal finding of Exp. V, namely that incidental recall when a clear picture is preceded by a blurred version of the same picture surpasses incidental recall when a clear picture is preceded by an unrelated blurred picture. The effect is, however, of comparable strength whether or not *S* knows in advance that the clear picture will depict the same object as the blurred picture. There is, therefore, no support for the view that preparatory mobilization of attention before the onset of the clear slide is responsible for the effect.

DISCUSSION

The findings of all six experiments concur with the hypothesis that information is more likely to be retained when its initial presentation relieves subjective uncertainty. The effect failed to appear when the blurred slide did not induce uncertainty or when the clear slide did not contain information that could reduce the uncertainty. As in the previous experiments with Turkish words (Berlyne et al., 1968), both the phase of uncertainty induction and subsequent phase of uncertainty reduction were found to be necessary. The effect can be interpreted as a matter of reinforcement due to removal of an aversive condition (satisfaction of epistemic curiosity). Alternatively, it may be asserted that subjective uncertainty makes *S* especially attentive to information that does away with the uncertainty. As pointed out elsewhere (Berylne, 1969, 1970a), psychologists use the word "attention" in several different senses, some of which make it synonymous with "reinforcement" (Berlyne, 1967, 1970b). Consequently, explanations of our findings in terms of epistemic curiosity and in terms of attention are, according to some usages, formally equivalent. Both kinds of explanation are compatible with the possibility that subjective uncertainty facilitates learning by promoting verbal, receptor-adjusting, postural, or other "mathemagenic" responses (Rothkopf, 1965), which facilitate the intake of information.

At least one kind of explanation invoking attention appears to be ruled out by Exp. VI. It was found that prior exposure to a related blurred picture facilitated recall even when *S* had no way of knowing whether the corresponding clear picture would follow. The effect does not therefore depend on attention in the sense of anticipatory mobilization of information-assimilating functions.

Data obtained by Parlow (1970; see Parlow & Berlyne, 1971), using the Turkish-word paired-associate technique of Berlyne et al. (1968), suggest that *S*'s guess may facilitate recall by serving as a mediator. After being given the stimulus term, *S* recalled a guess made by himself better than a guess made by a yoked-control *S* (presented as a "possible English translation, which may or might not be correct [Parlow, 1970, p. 36]"). And the response term was more readily recalled in the presence of the stimulus term and *S*'s own guess than in the presence of a stimulus term and a yoked-control *S*'s guess. Free associates did not function in this way; they were recalled more often than arbitrarily given words in response to the stimulus term, but, when they were paired with the stimulus term, they did not facilitate recall of the response term. In the present experiments, however, guesses could not serve as mediators, since there were no stimulus terms. But our *S*s invariably reported, when questioned subsequently, that they had tried to guess the identities of the objects depicted in some or all of the blurred slides. It is conceivable, therefore, that recall of incorrect guesses somehow facilitated recall of objects figuring in related clear slides.

Since subjective uncertainty induced by exposure to blurred pictures has been found to heighten arousal (Berlyne & Borsa, 1968), as have other conditions conducive to uncertainty (Berlyne, 1971), the effect investigated in these experiments might represent reinforcement through reduction of arousal from an aversive level. This is one of two mechanisms of reinforcement suggested by available evidence (Berlyne, 1967, 1971). The epistemic-curiosity explanation can, however, stand or fall by itself, whether or not fluctuations in arousal turn out to play a part.

REFERENCES

Berlyne, D. E. A theory of human curiosity. *British Journal of Psychology*, 1954, **45**, 180–191.

Berlyne, D. E. *Structure and direction in thinking.* New York: Wiley, 1965.

Berlyne, D. E. Arousal and reinforcement. *Nebraska symposium on motivation*, 1967, **15**, 1–110.

Berlyne, D. E. The development of the concept of attention in psychology. In C. R. Evans & T. Mulholland (Eds.), *Attention in neurophysiology.* London: Butterworths, 1969.

Berlyne, D. E. Attention as a problem in behavior theory. In D. Mostofsky (Ed.), *Attention: Contemporary theory and analysis.* New York: Appleton-Century-Crofts, 1970. (a)

Berlyne, D. E. Motivational problems. In J. Linhart (Ed.), *Proceedings of the international conference on psychology of human learning.* Prague: Institute of Psychology, Czechoslovak Academy of Sciences, 1970. (b)

Berlyne, D. E. *Aesthetics and psychobiology.* New York: Appleton-Century-Crofts, 1971.

Berlyne, D. E., & Borsa, D. M. Uncertainty and the orientation reaction. *Perception & Psychophysics*, 1968, **3**, 77–79.

Berlyne, D. E., Carey, S. T., Lazare, S. A., Parlow, J., & Tiberius, R. Effects of prior guessing on intentional and incidental paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 750–759.

Dunnett, C. W. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 1955, **50**, 1096–1121.

Nicki, R. M. The reinforcing effect of uncertainty reduction on a human operant. Unpublished doctoral dissertation, University of Toronto, 1968.

Nicki, R. M. The reinforcing effect of uncertainty reduction on a human operant. *Canadian Journal of Psychology*, 1970, **24**, 389–400.

Parlow, J. Effects of prior guessing on immediate recall of incidentally-learned verbal associations. Unpublished doctoral dissertation, University of Toronto, 1970.

Parlow, J., & Berlyne, D. E. The effect of prior guessing on incidental learning of verbal associations. *Journal of Structural Learning*, 1971, **4**, 55–65.

Rothkopf, E. Z. Some theoretical and experimental approaches to problems in written instruction. In J. D. Krumboltz (Ed.), *Learning and the educational process.* Chicago: Rand McNally, 1965.

(Received January 5, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 49–53

PERCENTAGE OF OCCURRENCE OF CORRECT RESPONSE AND IMPLICIT ASSOCIATIVE RESPONSES IN VERBAL DISCRIMINATION LEARNING

ROBERT W. NEWBY [1] AND ROBERT K. YOUNG

University of Texas at Austin

Each of 80 *S*s learned a verbal-discrimination (VD) list for six trials. Two levels of percentage of occurrence of the correct response (100% or 50%) were combined factorially with four conditions. In three of these conditions, background color was provided which represented the implicit associative response to the correct item 100%, 50% (50% to the incorrect item), or 0% (100% to the incorrect item) of the time. The fourth condition had no background color and served as a control. The data indicate that percentage of occurrence of the correct response was an important variable with performance being reduced considerably at 50% occurrence. In addition, when background color was associated with the incorrect response, performance was retarded under the 50% correct response occurrence condition. The results were interpreted primarily in terms of the frequency theory of VD learning.

The frequency theory of verbal discrimination (VD) learning (Ekstrand, Wallace, & Underwood, 1966) states that *S*s learn a VD list by means of Rule 1: Always choose the most frequent item. This rule is followed by *S* unless otherwise instructed (Underwood, Jesse, & Ekstrand, 1964). One of the responses necessary for learning by means of Rule 1 is the rehearsal-of-the-correct-alternative response (RCR) (Ekstrand et al., 1966). In those cases where *S* has responded with the correct (C) item, the RCR adds one frequency unit to the C item. However, when *S* has responded with the incorrect (I) item, the RCR, by adding a frequency unit to the C item, evens the frequency ratio between the C and I items. In this manner, learning can proceed by use of Rule 1 since the frequency ratio can never favor the I item.

However, Ekstrand et al. (1966) point out that the RCR is really necessary only in those instances where *S* has responded incorrectly with the I item. Therefore, it is entirely possible that learning could occur in a VD task without the RCR. If this is so, then in contrast to predictions from frequency theory, inhibition of the RCR would not lead to any decrement in performance.

If the RCR is to be inhibited, *S* must not be told which item is correct. To accomplish this, percentage of occurrence of correct alternative (analogous to percent occurrence of response members [% ORM] in paired-associate [PA] learning) was selected as a variable. ORM has been demonstrated to be inversely related to difficulty of learning in PA tasks (e.g., Goss, Morgan, & Golin, 1959; Schulz & Runquist, 1960) even though differences in rate of learning between highest and lowest percentages have been relatively small.

If the RCR is an important response in VD learning, a reduction in ORM should result in competition between the C and I items. Under reduced ORM the RCR can be made only with a certain percentage of pairs, and when it is not made following an incorrect response, frequency then becomes an unreliable basis for selecting the correct response. The reason for this is that incorrect alternatives sometimes will have a higher frequency than correct alternatives. This should not be the case at 100% ORM since the RCR is made on every trial.

Underwood (1965) found that sensory impression words failed to elicit IARs (implicit associative responses) in a free-recall task. However, his *S*s were not in-

---

[1] Requests for reprints should be sent to Robert W. Newby, Department of Psychology, University of Texas at Austin, Austin, Texas 78712.

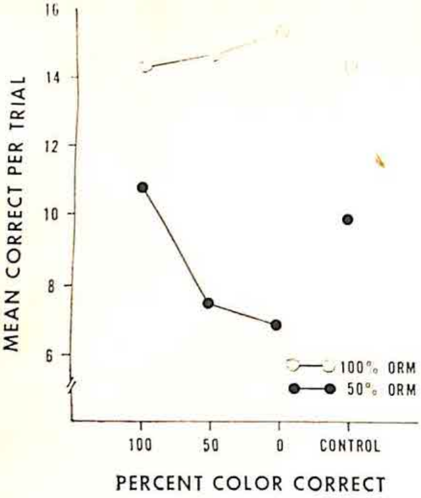


FIG. 1. Mean correct per trial as a function of % ORM and percent color correct.

formed of the relationship between the sensory impression words. In subsequent research, Wood and Underwood (1967) and Wood (1968) using a free-recall paradigm demonstrated that color-sensory impression words elicited IARs if *S*s were given a cue regarding the relationship between the words.

The present experiment was designed to determine whether color-sensory impression words elicit IARs when appropriate cues are present. Frequency theory would predict that if color-sensory impression words elicit IARs, they should facilitate performance when the IARs are associated with C items. When associated with I items or with both C and I items as in a mixed list, there should be an inhibition in performance. If the color-sensory impressions are not elicited as IARs, there should be no differences between groups.

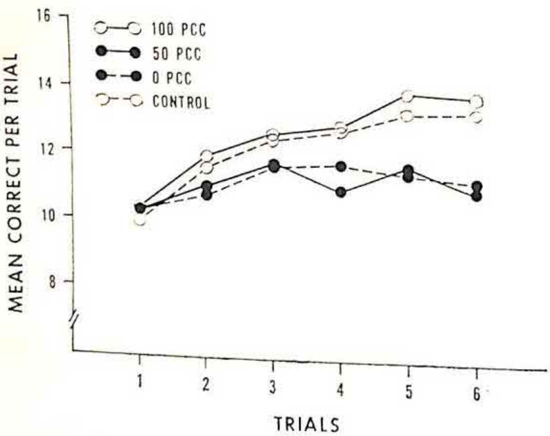


FIG. 2. Mean correct per trial as a function of trials and percent color correct.

The present experiment was designed to test two predictions: A reduction in ORM will inhibit the RCR resulting in a reduction in performance; color-sensory impression words will, when cued, elicit IARs which will facilitate performance when associated with C items and will inhibit performance when associated with I items or with both C and I items.

METHOD

Design.—Two levels of ORM were selected: 100% and 50%. At 100% ORM, the C item for each pair was underlined on every study trial. At the 50% ORM level, two sublists of eight pairs each were randomly selected. For Sublist O the C items were underlined on odd study trials and not underlined on even study trials, and for Sublist E the C items were underlined on even trials but not on odd.

Three levels of percent color correct (PCC) plus a control were selected. PCC indicated the percentage of the time that the color of slide was the color-sensory impression of the C item. The three levels were 100, 50, and 0 PCC. Thus, for example, in the 50 PCC group for each pair the color of the slide was the color-sensory impression of the C item on half the trials and of the I item on the other half of the trials. The same background color used in a slide was presented on both the study trial and the immediately following test trial for all PCC groups. The ORM and PCC variables were combined to yield a 2 × 4 factorial design ($n = 10$ per cell).

The ORM and PCC variables were manipulated orthogonally such that within each of the E and O sublists in the 50 PCC group at 50% ORM, the background colors were appropriate for the C items for half the pairs and for the I items for the other half of the pairs. The slides used for the Control group had no background color.

Procedure.—The study-test method was used in the experiment; and six test trials on the VD task, each preceded by a study trial, were given. A list of 32 nouns which elicit color-sensory impressions was selected from Underwood and Richardson (1956). The words were paired randomly with the restrictions that two words with the same color-sense impression were not paired together, and that the pairing was balanced over the four possible colors. Each *S* received the same 16-pair list with the exception of change of slide color depending on experimental condition.

The list was presented by means of a Kodak Carousal 650 projector with an attached timer at a 3-sec. rate with a 12-sec. intertrial interval for both study and test trials. The color was added to the slides by means of transparent plastic of the desired color. The four colors were red, green, yellow, and white. The two Control groups had no color added to their slides. A total of 80 *S*s from introductory psychology classes were run in groups of 5 and were randomly assigned to each of the eight experimental conditions. The *S*s were instructed to write the correct response for each pair presented in a six-page test booklet but were not forced to make a choice. After each test trial the page was turned and placed face down on the table in front of *S*. No *S* had previously served in a learning experiment, and no *S* was dropped from the experiment.

RESULTS

Analysis of the data revealed a significant effect for ORM with F (1, 72) = 174.43, $p < .001$. It can be seen in Fig. 1 that change of ORM from 100% to 50% resulted in a reduction in performance. Also significant was the ORM × PCC interaction with F (3, 72) = 6.75, $p < .01$. This interaction is also presented in Fig. 1. It appears that the PCC variable was not related to performance at the 100% ORM level while it was at the 50% ORM level. This was substantiated by separate (nonorthogonal) analyses performed at both levels of ORM. At the 100% ORM level, PCC was not significant, F (3, 36) = 2.09, $p > .05$, but at the 50% ORM level, PCC was significant with F (3, 36) = 5.14, $p < .01$. In addition, trials was also significant with F (5, 360) = 23.42, $p < .001$, indicating that performance improved over the six trials of the VD task.

There was also a significant Trials × PCC interaction, F (15, 360) = 2.51, $p < .01$. As can be seen from Fig. 2, this interaction resulted from the fact that the performance of the 100 PCC and Control groups improved at a more rapid rate than did the 50 and 0 PCC groups.

Half the correct responses were underlined on each study trial. That is, one half of the pairs (O sublist) had their respective correct responses indicated on Trials 1, 3, and 5, while the correct responses for the other half of the pairs (E sublist) were shown on Trials 2, 4, and 6. Thus an additional variable, sublist, can be added to the analysis. These data are presented in Fig. 3. It can be seen from Fig. 3 that underlining the correct response during a study trial increased performance on the subsequent test trial and that failure to underline the correct response decreased performance on the subsequent test trial. The Trials × ORM × Sublists interaction, presented in Fig. 3, was highly significant, F (5, 360) = 11.17, $p < .001$. (In this case the 100% ORM groups serve as a control for differential sublist difficulty.) From Fig. 3 it can be seen that much of the variability of this interaction was specifically related to differential performance on the test trial after the correct member of the pair was or was not indicated on the immediately preceding study trial. This is shown by the Sublists × Odd versus Even Trials × ORM interaction (not orthogonal to the previous interaction), F (1, 360) = 48.38, $p < .001$. Finally, the Trials × PCC × Sublists interaction was also significant, F (15, 360) = 3.20, $p < .001$. Specifically, this latter interaction was obtained because the alternating character of performance shown in Fig. 3 was missing in the 50 PCC group of the 50% ORM condition, while performance in the other three 50% ORM groups was virtually identical to that presented in Fig. 3.

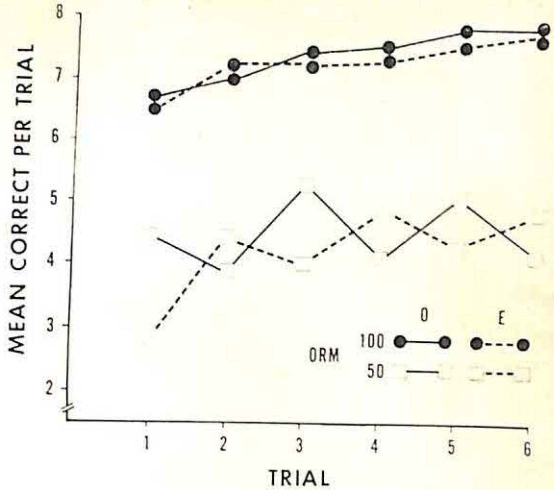


FIG. 3. Mean correct per trial for sublists on which correct response is underlined on odd trials (O sublist) and even trials (E sublist).

DISCUSSION

Ekstrand et al. (1966) seemingly give the RCR a position of secondary importance in their discussion of frequency theory. According to the authors, the main function of the

RCR is to even the frequency ratio of the pairs after a Pronunciation Response (PR) has been made to the I item. The present results indicate, however, that the RCR plays an important role such that when the RCR is inhibited, performance is reduced. This apparently indicates that unless *S* is allowed to make an RCR on every trial for each pair, the learning of a VD list becomes more difficult.

It will be recalled that *S*s were encouraged to respond during test trials but were not required to do so. Thus on some trials neither the correct nor incorrect alternative was chosen, and performance was probably lower than would have been the case if *S*s had been required to guess. However, a reliable difference between the 50% and 100% ORM conditions is found when a forced-choice procedure is used (Newby, 1972). In addition, the alternating character of performance shown in Fig. 3 is maintained under a forced-choice procedure (Newby, 1972) and also appears to be a replicable phenomenon.

The increasing and decreasing performance which paralleled the inclusion and omission of a line under the C item in the 50% ORM conditions can be accounted for by frequency theory. Assume that during study trials *S*s attend primarily to those pairs in which the C item is underlined. If such is the case, the frequency units accruing to the pairs in which the C item is *not* underlined must be approximately the same for both C and I items. On study trials the frequency difference between the C and I items remains constant for these pairs but frequency units increase for both C and I items, and the same C–I frequency difference yields a more difficult discrimination resulting in a decrement in test-trial performance. On the next study trial, the C item is underlined, differential numbers of frequency units accrue to the C and I items, and subsequent test-trial performance increases.

With the reduction of ORM, the increased difficulty of the task forced *S*s to look for other cues to aid in learning the VD task. This resulted in making *S*s aware that slide color could act as a cue for responding. That is, it seems *S*s cued themselves to the IARs provided by slide color. This supports the findings of Wood and Underwood (1967) and Wood (1968) who demonstrated that *S*s use sensory impressions as IARs when cued to do so.

Underwood (1965) found that sensory impressions did not lead to greater recall in a free-recall study. The nonsignificant difference found at the 100% ORM level between the four PCC groups supports his findings. However, at the 50% ORM level the groups differed, with the direction of the differences being consistent with predictions from frequency theory. That is, performance was facilitated when IARs were associated with C items, inhibited when associated with I or both C and I items.

An alternate interpretation of the effects of the PCC variable is suggested by research on the effects of imagery. That is, *S*s instructed to image the correct member of a VD pair learn faster than *S*s instructed to image both members of the pair (Rowe & Paivio, 1971). It may be that *S*s in the 50% ORM conditions used the background color as a cue to image one or both of the members of the pair. When the image item was a C item, performance was slightly facilitated, but when I items or both C and I items (as in the 50 PCC group) were imaged, performance was somewhat impaired.

From Fig. 2, it is apparent that the 100 PCC and Control groups improved in their performance over trials. In neither group is any response made to the I item except the pronunciation response (Ekstrand et al., 1966) which is made by chance on the first few trials. However, the 50 and 0 PCC groups reached their point of maximum performance by Trial 3, and their performance did not increase thereafter. This was so because frequency was not a reliable cue, yet *S* continued to use Rule 1. In these groups, frequency tended to favor the I item because of the IARs provided by the slide color or, alternately, because *S*s failed to image only the C item.

In summary, the following results were obtained. The RCR is a significant part of VD learning and its inhibition results in a marked decrement in performance. In addition, color-sensory impression words elicit IARs when *S* is cued in some manner, i.e., a reduction in ORM, but not when *S* does not need or receive a cue. Also when frequency is not a reliable cue as in the 0 and 50 PCC groups of the 50% ORM conditions, performance increases relatively slowly. The results can be taken as supportive of the frequency theory as proposed by Ekstrand et al. (1966).

REFERENCES

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

Goss, A. E., Morgan, C. H., & Golin, S. J. Paired-associate learning as a function of percentage of occurrence of response members (reinforcement). *Journal of Experimental Psychology*, 1959, **57**, 91–104.

Newby, R. W. Effects of percentage of occurrence of response members and list type on verbal discrimination transfer. Unpublished doctoral dissertation, University of Texas, 1972.

Rowe, E. J., & Paivio, A. Imagery and repetition instructions in verbal discrimination and incidental paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 668–672.

Schulz, R. W., & Runquist, W. N. Learning and retention of paired adjectives as a function of percentage occurrence of response members. *Journal of Experimental Psychology*, 1960, **59**, 409–413.

Underwood, B. J. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, 1965, **70**, 122–129.

Underwood, B. J., Jesse, F., & Ekstrand, B. R. Knowledge of rights and wrongs in verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 183–186.

Underwood, B. J., & Richardson, J. Some verbal materials for the study of concept formation. *Psychological Bulletin*, 1956, **53**, 84–95.

Wood, G. Implicit responses and conceptual similarity: A repetition. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 838–840.

Wood, G., & Underwood, B. J. Implicit response and conceptual similarity. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 1–10.

(Received January 10, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 54–59

CONCEPT ATTAINMENT AS A FUNCTION OF MOTIVATION AND TASK COMPLEXITY[1]

PATRICK R. LAUGHLIN,[2] RICHARD E. CHENOWETH,
BARBARA B. FARRELL, AND JOSEPH E. McGRATH

University of Illinois

In order to test the predicted interactive effects of motivation and task complexity, 96 college students solved three concept-attainment problems in a 2 × 3 × 2 × 3 factorial design with the variables: (*a*) stress (experimental session videotaped or not videotaped), (*b*) number of attributes in the concept universe (four, six, or eight), (*c*) number of values per attribute (two or three), and (*d*) successive problems (three of the same type per *S*). Pulse rates were taken five times as a physiological measure of the induced stress. Major results were the following: (*a*) higher pulse rates and a higher proportion of both untenable inference and untenable memory hypotheses for the video condition; (*b*) an interaction between stress, attributes, and values for focusing strategy, indicating a detrimental effect of stress for the most complex trivalued eight-attribute concept universe; (*c*) a linear increase in number of instances to solution with increasing attributes; (*d*) no main effect of number of values, but an Attributes × Values interaction for untenable memory hypotheses; (*e*) improvement over successive problems only in the most difficult eight-attribute concept universes.

The primary purpose of the following experiment was to assess the effect of motivation, as manipulated by videotaping the experimental session, upon concept attainment. This videotaping may be considered a complex motivational manipulation consisting of the stress resulting from the presence of a cameraman in addition to *E*, the evaluative, "dramaturgical" stress involved in being videotaped, and the implication of a permanent record of performance in the videotape itself. In his theory of social facilitation, Zajonc (1965) has proposed that initial learning of an activity is hindered by the presence of others, while performance of a well-learned activity is facilitated by the presence of others. Although concept attainment involves the performance of previously well-learned responses such as those involved in the perception of the stimulus display and comprehension of the experimental instructions, *S* must learn what attributes and/or rules have been designated relevant and perhaps learn a general strategy of solving the class of such problems; hence a main effect of motivation was predicted, with poorer performance for the videotaping condition. As a physiological check upon the effectiveness of the manipulation, pulse rates were taken at five points during the experiment. A main effect indicating higher pulse rates for the video conditions was expected.

The second purpose was to assess possible interactions between motivation and concept complexity, as the social facilitation theory of Zajonc (1965), Hebb-Malmo arousal theory (Cofer & Appley, 1964), and Hull-Spence drive theory (Logan, 1959) would all predict that performance under the video condition should become increasingly poorer relative to performance under the nonvideo condition with increasing concept complexity. As a test of this prediction, concept complexity was orthogonally varied by concept universes consisting of four, six, or eight total attributes, with two or three values per attribute.

---

[1] This research was supported in part by a grant from the University of Illinois Research Board (P. R. L., principal investigator) and in part by Air Force Office of Scientific Research Grant AF 1161-67 (J.E.M., principal investigator). Appreciation is expressed to Peter Wallner and James J. Farrell for serving as cameramen in the video conditions.

[2] Requests for reprints should be sent to Patrick R. Laughlin, Department of Psychology, University of Illinois, Psychology Building, Champaign, Illinois 61820.

This also enabled a determination of the main effects of these two independent manipulations of concept complexity and their interaction. The main effect of values was of particular interest, since no previous research with a selection paradigm has varied the number of values per attribute for a constant number of attributes and a positive focus instance (Miller, 1971). Since the number of possible hypotheses with a positive focus is determined solely by the number of attributes, irrespective of the number of values per attribute, the main effect of values was predicted to be nonsignificant.

METHOD

Design and subjects.—The experimental design was a $2 \times 3 \times 2 \times 3$ factorial with repeated measures on the last variable. The variables were the following: (*a*) stress (experimental session videotaped or not videotaped), (*b*) number of attributes (four, six, or eight), (*c*) number of values per attribute (two or three), (*d*) successive problems (three of the same type for each *S*). The *S*s were 48 male and 48 female introductory psychology students at the University of Illinois, fulfilling course requirements for experimental participation. Four of each sex were randomly assigned to each of the 12 between-*S*s conditions of the design. One male and one female *E* (the second and third authors) each ran 2 male and 2 female *S*s within each condition, so both the sex of *S* and the sex of *E* were completely orthogonal in the design.

As an additional control, the entire experiment was run in two successive blocks of 48 *S*s, with 1 *S* in each of the 48 unique stress–attribute–value–sex-of-*S*–sex-of-*E* conditions. Two sets of three problems each were randomly selected, and one of these problem sets was run in the first block of 48 *S*s and the other in the second block. Since all *S*s were randomly assigned to conditions (within the sex constraints), the experiment may be considered a randomized blocks design with two blocks, consisting of a composite of the two different problem sets, systematic effects such as learning or fatigue in the *E*s over time, and systematic *S* effects over time. The SS_{blocks} was subtracted from the $SS_{\text{error(between)}}$ for all dependent variables, resulting in a $MS_{\text{error(between)}}$ with 83 degrees of freedom rather than 84.

Procedure and instructions.—The concept instances were generated by a variation of Huttenlocher's (1962) selection procedure. The *S* sat at a table before two rows of four, six, or eight 10.16 × 15.24 cm. cards. The top row of cards consisted simply of the numbers 1–4, 1–6, or 1–8, and served to identify the attributes of the cards in the second row by ordinal position. Each attribute (ordinal position) had the values of a plus or minus sign in the bivalued concept universes, and a plus, minus, or zero sign in the trivalued universes. The second row consisted of four, six, or eight plus, minus, or zero signs according to the particular concept universe and constituted the known initial positive focus or instance of the concept. These signs were all lettered in green ink, so the given positive focus instance was referred to as the "green row." Depending upon the particular number of attributes and values in the concept universe, there were two or three piles of four, six, or eight plus, minus, and zero signs lettered in red ink on 10.16 × 15.24 cm. cards. The *S* generated concept instances by placing whatever one of these red signs he desired under each ordinal position or attribute. When *S* had generated a concept instance, or "red row," *E* said yes or no to indicate whether it was a positive or negative instance of the concept. The *S* then stated one (only) verbal hypothesis indicating what he considered the correct concept to be. The *E* then said yes or no to indicate whether or not the hypothesis was correct. A correct hypothesis solved the problem, while if the hypothesis was not correct *S* generated a second instance by arranging the appropriate signs in each ordinal position (attribute) of the red row. This cycle of instance generation, feedback on the instance, one hypothesis, and feedback on the hypothesis continued until *S* verbalized the correct hypothesis, which was the criterion of successful conceptual performance.

The problems were conjunctive concepts with two relevant attributes (e.g., "one plus and four zero"). Two sets of three problems each were randomly selected within each of the six attribute-value concept universes, and the initial instances (green rows) were randomly selected from the subset of possible positive instances for each concept. The same problems for a given concept universe were used for both stress conditions.

The instructions thoroughly explained with an example the nature of the task and the conjunctive rule, and explained the procedure just indicated. The *S* was informed that he would have three, two-attribute conjunctive problems and that the correct values of each attribute were on the given initial positive instance (green row). It was emphasized that the problems were to be solved in as few instance generations as possible, regardless of time.

In the video conditions, the cameraman (always male) began videotaping *S*'s performance after the conclusion of the instructions. The *S*s were simply informed that the session was being videotaped in order to obtain a record of performance. In all conditions, pulse rates were taken at the radial artery on the wrist five separate times by *E*. Pulse rates were measured by the number of seconds required for 20 beats rather than the usual method of counting beats within a fixed time interval, as this method requires *E* only to count 20 beats, then stop his watch and record, rather than to count beats while watching time. This measure was later converted to beats per minute.

Summary sequence of events.—In summary, the sequence of events was as follows: (*a*) *S* enters and

TABLE 1

Mean Number of Instances to Solution, Proportion of Untenable Hypotheses, and Focusing Strategy, for Three Problems and Totals over Three Problems

No. of attributes	No. of values	Problem	Instances to solution		Untenable hypotheses		Focusing strategy	
			Video	No video	Video	No video	Video	No video
4	2	1	2.50	2.13	.15	.10	.18	.30
		2	4.38	2.88	.31	.18	.16	.34
		3	3.88	3.00	.27	.05	.36	.40
		Total	10.75	8.00	.72	.33	.70	1.04
	3	1	2.50	2.75	.16	.00	.25	.37
		2	2.88	2.63	.14	.03	.29	.19
		3	3.13	3.13	.19	.06	.45	.15
		Total	8.50	8.50	.49	.09	.99	.70
6	2	1	5.25	4.38	.24	.10	.31	.52
		2	3.75	4.38	.05	.08	.51	.54
		3	4.13	5.50	.17	.19	.63	.48
		Total	13.13	14.25	.47	.37	1.45	1.53
	3	1	3.25	3.13	.16	.02	.40	.50
		2	5.13	5.25	.19	.16	.25	.29
		3	6.63	4.38	.26	.10	.20	.43
		Total	15.00	12.75	.61	.27	.85	1.22
8	2	1	9.88	6.63	.23	.15	.44	.33
		2	5.75	7.13	.08	.11	.47	.39
		3	4.62	5.63	.03	.04	.65	.23
		Total	20.25	19.38	.34	.30	1.56	.95
	3	1	9.00	4.75	.34	.09	.18	.43
		2	3.88	3.38	.19	.08	.29	.58
		3	3.38	4.25	.09	.19	.32	.47
		Total	16.25	12.38	.62	.36	.79	1.48

is seated; (*b*) first pulse rate is taken; (*c*) concept attainment instructions are given; (*d*) videotaping begins in stress conditions; (*e*) second pulse rate is taken; (*f*) first concept problem is given; (*g*) third pulse rate is taken; (*h*) second concept problem is given; (*i*) fourth pulse rate is taken; (*j*) third concept problem is given; (*k*) videotaping ends; (*l*) fifth pulse rate is taken; and (*m*) debriefing occurs. The video equipment and cameraman were in the room from the beginning to the end of the experiment, and the cameraman simply stood at the camera until he began videotaping. Both camera and cameraman were absent entirely in the nonvideo conditions.

Results

Means for the three major dependent variables of number of instances to solution, proportion of untenable hypotheses, and focusing strategy are given in Table 1.

Pulse rate.—The main effect of stress was highly significant, F (1, 83) = 12.61, $p <$ .001, with higher pulse rates for the video condition. Thus, as a manipulation check, this physiological measure indicates that the motivational manipulation was successful in inducing stress experimentally, apart from the effect of this induced stress upon performance. Neither the main effect of attributes nor values, nor any of their interactions, were significant, indicating that concept complexity per se was not a motivational manipulation. The main effect of the five successive pulse measurements was significant, F (4, 336) = 10.14, $p <$ .001, as was the Successive Pulse Measurements × Stress interaction, F (4, 336) = 4.98, $p <$ .01. The simple main effects of successive measurements for levels of stress indicated significant differences between the five pulse measurements for the video condition, F (4, 336) = 13.87, $p <$ .001, but no differences for the nonvideo condition. Within the video condition, Newman-Keuls comparisons indicated a highly significant ($p <$ .001) increase from the first to the second measurement of pulse rate as a function of the concept attainment instructions and beginning of videotaping. Since no comparable effect occurred in the nonvideo condition, this is a further indication of the effectiveness of the video manipulation in inducing stress or arousal. In comparison to this peak level of arousal on the second pulse rate, there was a significant decrease on each of the three subsequent pulse measurements, but the arousal level on the fifth measurement (after the videotaping ended) was still significantly higher than the initial level.

Number of instances to solution.—The main effect of stress was not significant, $F(1, 83) = 1.59$. The main effect of attributes was highly significant, $F(2, 83) = 17.15$, $p < .001$. Means for concept universes with four, six, and eight attributes were 8.94, 13.78, and 17.06, respectively, and by Newman-Keuls comparisons all three pairs of means differed from each other ($p < .001$). Thus, performance decreased linearly with increasing numbers of attributes in the concept universe. Neither the main effect of values nor any of the interactions were significant.

The main effect of successive problems was not significant, but there was a significant Problems × Attributes interaction, $F(4, 168) = 6.52$, $p < .001$. The simple main effects of problems for levels of attributes indicated significant change over problems only for eight attributes, $F(2, 168) = 11.43$, $p < .001$. Means within the eight-attribute concept universe were 7.56, 5.03, and 4.47 for Problems 1, 2, and 3, indicating significant improvement from Problem 1 to Problems 2 ($p < .001$) and 3 ($p < .001$), with no difference between Problems 2 and 3.

Proportion of untenable hypotheses.—Each hypothesis was classified as tenable or untenable, or logically consistent or inconsistent with available information. With the conjunctive rule and two relevant attributes there were two types of untenable hypotheses: (*a*) Both values of the hypothesis appeared on a previous negative instance; (*b*) either of the other value(s) of either attribute of the hypothesis appeared on a previous positive instance (e.g., the hypothesis "one plus and two minus" would be untenable if either a one-minus-or-zero or two-plus-or-zero instance had been positive). Hence, a hypothesis had to be consistent with both the current instance and all previous instances in order to be tenable, so the measure subsumed both the "perceptual-inference errors" (responses inconsistent with the current instance) and "memory errors" (responses consistent with the current instance, but inconsistent with any previous instance) of Cahill and Hovland (1960). The number of untenable hypotheses on each problem was divided by the total number of instances to solution on that problem to obtain the proportion of untenable hypotheses.

The main effect of stress was significant, $F(1, 83) = 12.12$, $p < .001$, with more untenable hypotheses for the video conditions. Neither the main effects of attributes nor values nor any of the interactions between stress, attributes, and values were significant. Although the main effect of successive problems was not significant, there was a significant Problems × Attributes interaction, $F(4, 168) = 2.52$, $p < .05$. Only the simple main effect of problems for eight attributes was significant ($p < .05$), with means of .20, .11, and .08 for Problems 1, 2, and 3, indicating a higher proportion of untenable hypotheses for Problem 1 than Problem 3 ($p < .05$) and no difference between Problems 1 and 2, or 2 and 3. Thus, as with the number of instances to solution, there was improvement over successive problems only in the most difficult eight-attribute concept universe.

Proportion of untenable perceptual-inference hypotheses and proportion of untenable memory hypotheses.—As previously noted, the proportion of untenable hypotheses included both hypotheses that were inconsistent with the current instance and hypotheses that were inconsistent with any previous instance although consistent with the current instance. Accordingly, separate analysis of variance were conducted for each of these two types of untenable hypotheses. The results for these analyses were virtually indentical to those for the overall proportion of untenable hypotheses, with the single exception that the Attributes × Values interaction was significant for untenable memory hypotheses, $F(2, 83) = 3.62$, $p < .05$. Analyses of the simple main effects of attributes for levels of value indicated that there was no difference between four, six, and eight attributes for bivalued concept universes, while the simple main effect of attributes for trivalued concept universes was significant, F (2, 83) = 3.74, $p < .05$. Means for four, six, and eight attributes were .05, .13, and .14, indicating a greater proportion of untenable memory hypotheses for trivalued concept universes with six ($p < .05$) or eight ($p < .05$) attributes than four, with no difference between six and eight. Thus, there were more untenable memory hypotheses with increasing attributes for trivalued concept universes but no effect for bivalued concept universes.

In summary, the more detailed analyses of subtypes of untenable hypotheses gave no further information beyond the overall analyses, except for the Attribute × Value interaction for memory hypotheses. The video manipulation resulted in a significant increase in both inference and memory errors, while both the number of attributes and the number of values had no effect in themselves but an interactive effect on untenable memory hypotheses.

Focusing strategy.—Focusing strategy was scored according to two rules: Rule 1, in which each instance had to obtain information on a new attribute and Rule 2, in which the hypothesis had to be tenable. New information was obtained if the instance altered only one attribute not previously proven irrelevant (conservative focusing), or, if more than one attribute was altered (focus gambling), the instance was positive or the ambiguous information on a negative instance was correctly resolved on the next instance by altering only one attribute. Scoring rules for untenable hypotheses have been given above. The total number of instances that satisfied both of these two rules was divided by the total number of instances on the problem to obtain a continuous focusing score from .00 to 1.00.

None of the main effects of stress, attributes, or values, nor the three, two-way interactions were significant. However, there was a significant Stress × Attributes × Values interaction, F (2, 83) = 3.56, $p < .05$. Further analyses of this interaction indicated that the only significant simple interaction effect was Stress × Values for 8 Attributes, F (1, 83) = 6.42, $p < .025$. Means for the four stress-value conditions with eight attributes were video-two = .52, video-three = .26, nonvideo-two = .32, and nonvideo-three = .49. The simple, simple main effects for these four means indicated significantly more focusing for two-value than three-value concept universes for the video condition, F (1, 83) = 4.50, $p < .05$, with no significant difference between two and three values for the nonvideo condition. Thus stress resulted in less focusing only in the most complex eight-attribute, trivalued concept universe. The results for focusing strategy therefore supported the predicted Motivation × Concept Complexity interaction, but only at the highest level of concept complexity.

Neither the main effect of successive problems nor any of the interactions with problems were significant.

DISCUSSION

The major result of the experiment was poorer performance under the video condition for both overall untenable hypotheses and the two component measures of untenable inference hypotheses and untenable memory hypotheses. As stress increased there were a greater number of logical errors, both those in relation to the current instance and those in relation to previous instances. The results for inference errors are particularly important in indicating the detrimental effect of motivation or stress upon performance, since previous research has found very few inference errors in relation to the number of memory errors (Bourne, Goldstein, & Link, 1964; Cahill & Hovland, 1960). The failure to find a difference in number of instances to solution as a function of motivation may be interpreted as indicating the greater sensitivity of the measures of proportion of untenable hypotheses than the number of instances to solution, since the proportions control for lucky hypotheses before all hypotheses except the correct answer have been eliminated. A comparable result was obtained in another large factorial study (Laughlin, 1969).

Although the main effect of stress was not significant for focusing strategy, there was a significant Stress × Attributes × Values interaction for the most complex trivalued eight-attribute concept universe. Thus, the predicted Motivation × Concept Complexity interaction was obtained, but only at the highest level of concept complexity and for only one of the dependent variables. Since both this interaction and the main effect of stress were predicted from Zajonc's theory of social facilitation, both results support the theory. The interaction would also be predicted by both Hebb-Malmo arousal theory and Hull-Spence drive theory, so the results are equally supportive for these two theoretical positions.

There was a linear decrease in performance as measured by the number of instances to solution as a function of increasing attributes. This would be expected because the number of possible hypotheses increases directly with the number of attributes. However, there was no decrease in performance for either the measures of untenable hypotheses or focusing strategy, which indicates that the problems were solved equally logically at all levels of attributes. Again, this indicates the greater sensitivity of the measures of untenable hypotheses since they are based on proportions of tenable/untenable hypotheses rather than the total number of instance selections only. The results for the number of instance selections are in accord with a large amount of previous research in which the number of irrelevant attributes increased with a constant number of relevant attributes (see Laughlin, 1971, for references).

In contrast to the strong effects of increasing numbers of attributes on instances to solution, the main effect of two versus three values was not significant. This was expected because with a positive focus the number of possible hypotheses is completely determined by the number of attributes, irrespective of the number of values. In contrast, increasing values per attribute would be expected to decrease performance for either a negative focus or no-focus condition, in which the number of possible hypotheses is a function of both the number of attributes and the number of values, and such results were obtained by Battig and Bourne (1961) and Overstreet and Dunham (1969). However, even with the positive focus the number of values did result in poorer performance at the higher levels of numbers of attributes.

Finally, although the main effect of successive problems was nonsignificant for all dependent variables, there was a significant Problems × Attributes interaction for four different variables, indicating improvement over successive problems only for the most difficult eight-attribute concept universe. The general conclusion from both this and comparable previous results (Laughlin, 1969) seems to be that positive interproblem transfer occurs with a selection paradigm only at the higher levels of problem difficulty.

REFERENCES

Battig, W. E., & Bourne, L. E., Jr. Concept identification as a function of intra- and inter-dimension variation. *Journal of Experimental Psychology*, 1961, **61**, 329–333.

Bourne, L. E., Jr., Goldstein, S., & Link, W. E. Concept learning as a function of availability of previously presented information. *Journal of Experimental Psychology*, 1964, **67**, 439–448.

Cahill, H. A., & Hovland, C. I. The role of memory in the acquisition of concepts. *Journal of Experimental Psychology*, 1960, **59**, 137–144.

Cofer, C. N., & Appley, M. H. *Motivation: Theory and research.* New York: Wiley, 1964.

Huttenlocher, J. Effects of manipulation of attributes on efficiency of concept formation. *Psychological Reports*, 1962, **10**, 503–509.

Laughlin, P. R. Selection versus reception concept-attainment paradigms as a function of memory, concept rule, and concept universe. *Journal of Educational Psychology*, 1969, **60**, 267–273.

Laughlin, P. R. Concept attainment as a function of information specification and concept complexity. *Journal of Experimental Psychology*, 1971, **90**, 334–343.

Logan, F. A. The Hull-Spence approach. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. *General systematic formulations, learning, and special processes.* New York: McGraw-Hill, 1959.

Miller, L. A. Hypothesis analysis of conjunctive concept-learning situations. *Psychological Review*, 1971, **78**, 262–271.

Overstreet, J. D., & Dunham, J. L. Effect of number of values and irrelevant dimensions on dimension selection and associative learning in a multiple-concept problem. *Journal of Experimental Psychology*, 1969, **79**, 265–268.

Zajonc, R. B. Social facilitation. *Science*, 1965, **149**, 269–274.

(Received January 11, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 60–66

PARALLEL AND SERIAL PROCESSING IN TACHISTOSCOPIC RECOGNITION:

TWO MECHANISMS[1]

A. O. DICK[2]

University of Rochester

The existing data in tachistoscopic recognition suggest the existence of parallel and serial processes. Two experiments were carried out to examine the effects of visual noise masking on a parallel mechanism and postexposure cuing on a serial mechanism. Since both manipulations were found to be effective in the same experiment, it was concluded that two mechanisms are operative.

In a multi-item tachistoscopic display, accuracy is not consistent across the array of items. In the typical experiment, letters in peripheral retinal locations are reported more accurately than letters in more central positions. Woodworth (1938) suggested that the phenomenon was due to "spatial masking" created by adjacent letters.[3] The existence of performance inequalities within a display is well established, although the specific relation of position to accuracy varies from experiment to experiment. One implication of this result bears on the kinds of processing that underlie performance on tasks of this sort. Because simultaneously presented elements seem to interfere with each other, the inference must be that some parallel processes are taking place. It is doubtful if such effects are of retinal origin, since spatial masking does not depend on the physical spacing between items (e.g., Harcum, 1964). Parallel processes in visual analysis are to be expected, of course, since the input is parallel.

Just as the spatial masking effects imply parallel processing, many other data suggest that portions of visual analysis are done in a serial manner. Since the studies of Sperling (1960), the evidence showing a rapid decline in accuracy during the first .25–.5 sec. following termination of a visual display has been interpreted as support for a serial mechanism. That is, if all of the items were processed simultaneously, there would be no decline in accuracy as a function of delay as long as analysis of the items commenced immediately. Since a decline is obtained, however, it is more likely that this procedure reflects serial mechanisms. The act of responding, obviously, must be a serial process since language and verbal behaviors are organized in a serial manner.

In tachistoscopic recognition, input is parallel and output is serial. At some point the multiple parallel input must be converted into a sequential mode for multiple output. Since the position effect reflects parallel phenomena and postexposure cuing apparently reflects serial phenomena, it would seem consistent to suggest that two different mechanisms are involved. The relation of these two hypothetical mechanisms was explored in the present experiments. The logic of the experiments was as follows: Accuracy of report of part of a display should depend upon the total number of items available in the display. A visual noise mask can alter accuracy presumably by changing the number of available items. Thus, if a visual noise mask is added to one part of the display, accuracy should increase on other parts because the mask should reduce the number of available items and thus reduce the total amount of necessary processing. Further, the greater the number of unreported items masked, the higher should be the accuracy. This predicted outcome would support a parallel mechanism by way of altering the amount of spatial masking.

---

[1] This research was supported by National Science Foundation Grants GB 7848 and GB 17138 and United States Office of Education Grant OEG-2-710371B. Experiment I was reported at the Eastern Psychological Association Meetings, Atlantic City, 1970. Robin Loader carried out Exp. II.

[2] Requests for reprints should be sent to A. O. Dick, Department of Psychology and Center for Visual Science, University of Rochester, Rochester, New York 14627.

[3] Although this paper discusses problems associated with backward masking and spatial masking, no attempt is made to evaluate theories of masking, per se, or to integrate the phenomena. Rather, spatial masking is used as an example and backward masking as a technique. See Townsend, Taylor, and Brown (1971) for a brief review and further data on spatial masking.

By contrast, the use of a postexposure indicator yields evidence about decay characteristics of information in the visual system and serial properties of the system. Since this is postulated to be different from spatial masking, the amount of decay observed should not be altered by a reduction in the amount of spatial masking. In short, increasing the number of items masked should increase accuracy but delaying a report cue should decrease accuracy, and both effects should be observed simultaneously. These predictions were tested in two experiments in which three rows of letters were presented and *S* asked to report one of them. The specific hypotheses were tested by manipulating the number of rows masked, the delay of the mask, and the delay of the report cue.

EXPERIMENT I

Method

Subjects.—The *S*s were 24 summer students at Lake Forest College; all were volunteers.

Stimuli.—Three sets of 24 cards were prepared by placing Deca-Dry Lettering (Style #2736) on white cards. There were 12 letters per card with no repetitions; letters were randomly drawn with the requirement that no letter appear twice in the same sequence. The 12 letters were arranged in three rows of four with each letter subtending a visual angle of 27′ and the entire display subtended an angle of 2°30′ × 2°30′. Patterned masks consisting of cross-hatching were prepared with India ink and a Leroy pen. A template was used to draw the lines evenly spaced. When viewed in the tachistoscope, the grid subtended an angle of .5° × 3°40′. Two sets of 48 masks were prepared; each incorporated zero, one, or two grids in every possible combination.

Procedure.—The stimuli were shown in a Scientific Prototype three-field tachistoscope. The letters were presented in Channel 1 for 50 msec. and the masks in Channel 2 for 75 msec. The blank field was used to provide an adapting field (about 8 ftl.) containing a small black fixation point. The fixation point was centered midway in the middle row and was present at all times except when one of the other stimuli was shown. Under the no-mask condition, a blank card was flashed in place of the grid. This control is necessary because *S* can detect the switch from one field to another in the tachistoscope and this switch may have a disruptive effect on performance.

Design.—A partial-report procedure was used in which *S* was presented three rows of letters and asked to report part of the display, i.e., one of the rows. Coded auditory tones were used as report cues; a high tone (1,000 Hz.) indicated report of the top row, a middle tone (400 Hz.) report of the middle row, and a low tone (200 Hz.) report of the bottom row. The interval between the termination of the target and the initiation of the tone was either 0 or 200 msec. A visual mask was used and either followed the stimulus display immediately (0 msec.) or after a delay (200 msec.). The number of rows that were masked was also a variable with zero, one, and two of the three rows being masked. For a given row to be reported, e.g., top, this necessitates four masks: no mask (a blank card); one grid on the bottom row; one on the middle row; two grids, one each, on the middle and bottom rows. Thus, there were four variables involved: (*a*) delay of the report cue (tone), (*b*) delay of the mask, (*c*) row reported, and (*d*) row masked. This constitutes a total of 48 experimental conditions. Each *S* was tested on two randomized blocks of these 48 conditions, giving a total of 96 experimental trials. In addition, each *S* was given 24 free-recall practice trials prior to the experimental testing. In all cases, *S* reported the letters verbally.

Results and Discussion

The number of letters correctly reported was used as the dependent measure in a number of analyses. One overall analysis of variance (*S*s × Number of Masks × Delay of Tone × Delay of Mask × Row Reported) was carried out which showed (*a*) all main effects to be significant ($p < .001$) and (*b*) the interactions of Number of Masks × Delay of Mask, Number of Masks × Row Reported, and Delay of Mask × Row Reported (all with $p < .001$). A second set of analyses involved separating the data into three categories according to the number of rows masked. The mean squares for these three analyses are presented in Table 1: The characteristics of the overall analyses are preserved in Table 1 in that comparisons across a row yield an index of the relative contributions of each variable relative to the number of rows masked.

TABLE 1

Modified Source Tables for Experiments I and II: Entries Are Mean Squares

Source	Exp. I				Exp. II					
	Report unmasked				Report unmasked				Report masked	
	df	No mask	One mask	Two masks	*df*	No mask	One mask	Two masks	One mask	Two masks
Total	287	3.70	(4.15)	5.18	275	1.51	(1.79)	1.85	1.43	(1.26)
Delay of mask (DM)	1	.89	55.00	282.03	1	2.09	19.97	91.60	31.34	160.88
Error (DM × *S*s)	23	1.87	3.08	4.08	22	.94	.81	1.55	.95	1.76
Delay of Tone (DT)	1	8.68	6.67	34.03	1	13.04	16.35	1.92	21.48	.59
Error (DT × *S*s)	23	1.33	1.97	2.02	22	1.37	.50	.96	.81	.80
Row Reported (R)	2	61.56	101.72	54.01	2	4.22	10.77	8.11	8.91	2.30
ϕ_1 (M vs. T + B)	1	75.11	39.38	79.51	1	4.17	21.20	15.02	4.53	2.00
ϕ_2 (T vs. B)	1	48.00	164.06	28.52	1	4.26	.33	1.22	11.50	2.61
Error (R × *S*s)	46	5.50	3.26	4.82	44	.68	2.45	1.76	.68	.93
DM × DT	1	6.12	2.25	.17	1	2.45	3.67	8.00	20.38	.36
Errror (DM × DT × *S*s)	23	3.46	1.45	1.47	22	1.94	1.13	1.08	.86	.67
DM × R	2	4.35	2.94	15.17	2	2.09	.80	5.75	19.30	3.88
ϕ_1 (M vs. T + B)	1	8.51	1.06	25.00	1	1.04	.40	8.62	26.09	1.24
ϕ_2 (T vs. B)	1	.19	4.82	5.33	1	3.13	1.20	2.88	12.52	6.52
Error (DM × R × *S*s)	46	1.62	2.09	2.48	44	.97	1.11	1.01	.41	.65
DT × R	2	6.51	2.42	4.04	2	1.70	3.79	2.31	.97	.39
ϕ_1 (M vs. T + B)	1	.01	3.90	1.56	1	3.20	7.50	2.22	.18	.66
ϕ_2 (T vs. B)	1	13.02	.94	7.52	1	.20	.07	2.40	1.76	.13
Error (DT × R × *S*s)	46	1.88	1.48	1.82	44	1.23	.61	.91	.81	.67
DM × DT × R	2	1.12	3.54	5.51	2	1.93	2.16	.44	25.17	.19
ϕ_1 (M vs. T + B)	1	2.25	1.32	8.03	1	.18	4.31	.22	.26	.26
ϕ_2 (T vs. B)	1	.09	5.75	3.03	1	3.67	.03	.66	50.00	.13
Error (DM × DT × R × *S*s)	46	1.42	1.42	1.29	44	1.33	.94	.78	1.03	.59

Note.—T = top row, B = bottom row, M = middle row.

In Fig. 1 are shown the effects of the number of masks, the delay of the mask, and delay of the tone. From the figure it is clear that there is little difference between no mask and one mask on overall performance. Part of this lack of difference is due to the fact that the one-mask condition actually contains two conditions. For example, if the top row was reported, either the middle or the bottom row might be masked. There should be less facilitation, in this case, by masking the bottom row than by masking the middle. A more detailed analysis suggests that the single mask is more effective if it falls on an adjacent row than if it is removed by a row. Therefore, as displayed in Fig. 1, the condition with one mask underestimates the beneficial effect of masking.

The two-row masked condition shows clearly the benefit of the masking stimulus and the effects of delaying the cue. There is, however, a confound in this condition. Having two rows masked immediately constitutes a complete (but negative) visual cue, because *S*s were told that they would not have to report a row that was masked. This would be especially true for the case in which there were two masks occurring immediately after the display. The general effect of redundant cuing should be to reduce the magnitude of the decline in accuracy due to delaying the tone. Nevertheless, the data indicate that the tone has an effect since a delayed tone produced lower accuracy than an immediate tone (Table 1).

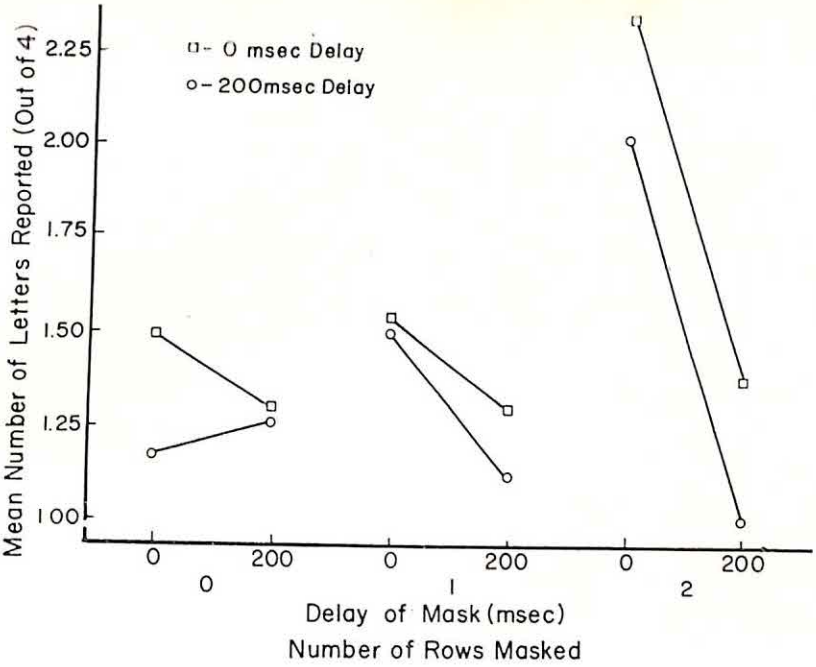


FIG. 1. Mean accuracy of report as a function of delay of the mask and number of rows masked. (Delay of the cue is the parameter. For the zero-mask condition, the mask delay refers to the delay in switching to a blank masking field.)

EXPERIMENT II

In Exp. I, the occurrence of the mask could have provided information about which row to report since *S*s were instructed that they would not have to report a masked row. This problem was eliminated in the second experiment by requiring report of all possible combinations, that is, sometimes *S* was asked to report a row that had been masked.

Method

Subjects.—The *S*s were 23 undergraduates at the University of Rochester taking an introductory course in psychology. Participation in the experiment fulfilled part of the class requirement. All *S*s were tested on a Snellin chart for visual acuity prior to participation in the experiment.

Stimuli.—The stimuli were some of those used in Exp. I.

Procedure.—The stimuli were presented in a Gerbrands three-field tachistoscope. The balance of the procedure was identical to Exp. I. The visual angle of the display was 3°20′ × 3°20′ shown at a background luminance of about 2 ftl. The exposure durations for both the mask and stimulus were 50 msec.

Design.—The four variables of Exp. I were factorially manipulated in the experiment. These consisted of (*a*) the seven different masks, (*b*) the three different rows that could be reported, (*c*) the delay of the report cue, and (*d*) the delay of the mask. Altogether, this gives 84 experimental combinations; the increase in conditions is due to the report of masked rows. To achieve this, 42 stimulus cards were presented twice; *S*s were cued to report different rows on two occasions, however. In addition, each *S* was given 12 practice trials before presentation of the experimental conditions. In all cases, *S*s reported the letters verbally.

Results and Discussion

As in Exp. I, two sets of analyses were applied to the number of letters correctly reported. The mean squares for the individual analyses on each of the masking conditions are presented in Table 1, the means are shown in Fig. 2 and Table 2. An analysis of variance applied to the data for unmasked row report showed the following effects to be significant: delay of the cue, delay of the mask, the number of rows masked, and the row reported ($p < .01$). In addition, virtually all of the interactions involving the row re-

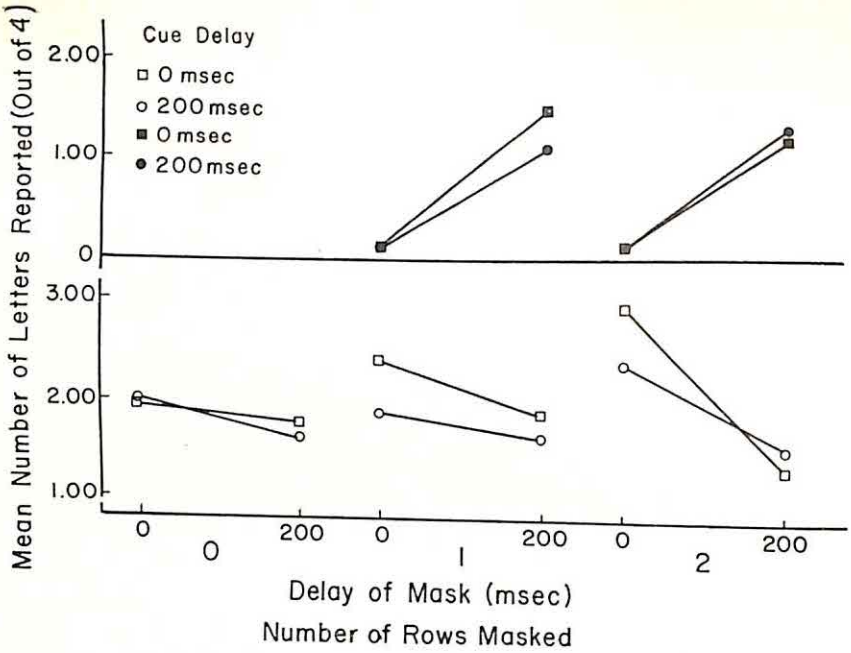


FIG. 2. Mean accuracy of report as a function of delay of the mask and number of rows masked. (Delay of the cue is the parameter. Filled symbols represent report of masked rows; open symbols report of unmasked rows. As in Exp. I, delay of the unmask condition refers to the delay of a blank field.)

ported were significant ($p < .01$) as in Exp. I. Most of these interactions are due to the fact that there is a bias toward the middle row where the fixation point was located and a slight bias away from the bottom row. Manipulations of the type of mask and mask delay dramatically change the relation of these biases. The data indicating the bias of row reported may seem to be a confounding result, but it is an important one to which the discussion will return. Before discussing the general implications, it is appropriate to consider some of the other effects in detail.

The greater the amount of the display that is masked, the higher the performance for report of unmasked rows. In addition, the mask is minimally effective when it is delayed by 200 msec.; in fact, performance is about the same for both masked and unmasked rows when the mask is delayed by 200 msec. The result that the number of rows masked changes performance has some implications for the type of processing that takes place. Clearly, processing cannot be carried out entirely in a serial manner; otherwise masking an unreported row would have no effect on a row that was reported. This result parallels that of Mewhort (1967) in which he showed that the familiarity of unreported material influences accuracy of reported material.

Delaying the report cue generally resulted in a reduction in accuracy; this effect was enhanced by masking one or two rows although the effect interacts with delay of mask. Without a mask, spatial interference appears to reduce the effects of decay. Since the magnitude of the decay effect increases as a result of masking, this suggests that decay is a postmasking phenomenon and is clearly consistent with the two-mechanism hypothesis.

CONCLUSIONS AND IMPLICATIONS

Taken together, the data on masking unreported rows and delay of the report cue suggest two different mechanisms. That is, the mask seems to affect the system at a point at which processing occurs in parallel and before complete identification of the stimulus has occurred. The report cue probably has its effect later in the processing sequence and in such a way that processing appears to be serial. The fact that an auditory signal yields evidence related to visual processing suggests that the locus of decay takes place relatively advanced in the system. In somewhat different terminology, the mask may affect preattentive processes, but the cue may affect attentive processes (Neisser, 1967).

The mechanisms might operate in the following way. A multielement input establishes an iconic memory. Information in iconic memory is not stored passively, but rather a number of operations are performed on its contents. (In this context, iconic memory describes poststimulus but preverbal analyses.) The analyses carried out at this point might include spatial position, shape, orientation, etc. This parallel system is connected to a sequential system in which verbal labels are attached. Although the visual analysis may be complete on an item, this does not automatically mean that the item will get transferred into short-term memory. The order of transfer is determined either by a self-induced set (e.g., the use of reading habits) or by experimental instructions. A report cue would be such an experimental instruction. If such a cue is delayed, *S* will begin to transfer items according to his own instructions, including a bias toward particular rows. When the cue arrives, he will have to switch attention on a certain proportion of the trials and alter the order of transfer. Furthermore, a switch in attention would require a finite amount of time.

TABLE 2

Mean Accuracy (out of 4) for the Various Conditions in Experiment II

Rows masked	Delay of mask	Delay of cue: 0 msec. T	0 msec. M	0 msec. B	200 msec. T	200 msec. M	200 msec. B
O	0	1.78	2.74	1.39	1.65	2.26	2.13
	200	1.39	1.91	2.04	1.87	1.70	1.44
T	0	.04	3.30	2.26	.09	2.17	1.87
	200	1.00	2.35	1.70	.87	1.96	1.13
M	0	2.70	.09	2.00	2.04	0.04	2.00
	200	1.70	2.52	1.22	1.44	2.26	1.83
B	0	2.00	2.44	.09	1.17	2.26	.09
	200	2.22	2.22	1.00	1.13	2.70	.30
T and M	0	.13	.04	3.26	.13	.30	2.48
	200	.87	2.44	1.17	.74	2.39	.96
T and B	0	.04	2.87	.04	.04	2.70	.00
	200	1.22	1.96	.56	1.13	2.30	1.00
M and B	0	2.70	.30	.22	2.04	.22	.17
	200	.96	1.22	1.04	1.39	1.78	1.09

Note.—Rows reported: T = top row, M = middle row, B = bottom row.

Several predictions are generated from the attention-switching hypothesis. First, decay would be observed when attention is focused on the wrong part of the stimulus and a switch would be required. Accuracy is lower because of (*a*) the time spent on the wrong portion of the display and (*b*) the time required for a switch in attention. Holding (1971) has reported data relevant to this prediction. He asked *S*s to guess which row they would be asked to report. Holding presented a three-row display and found that the accuracy of the guess about the row to be reported was inversely related to the amount of decay observed. If *S* correctly anticipated the correct row, no decay was observed. If *S* was off by one row, some decay was observed; if he was off by two rows, more decay was observed. The inequality of row accuracy in the present experiments is consistent with Holding's interpretation. Although the Row × Delay of Cue interaction is not significant, this is probably due to the small number of delays used and to the presence of the mask. The second prediction has to do with the switch itself. On a full-report trial, all of the items would be reported and hence a switch would not be needed. This suggests that accuracy in full report should be slightly higher than that of partial report. Taking output interference into account, Dick (1971) has shown that, indeed, full report is more accurate than partial report.

The results reported in this article and the interpretation given to them are in accord with other data (Eriksen & Colegate, 1971; Eriksen & Rohrbaugh, 1970). Eriksen and Colegate presented eight letters which were preceded by, simultaneous with, or followed by one or two indicators. Their results show that an indicator that precedes the display leads to better accuracy than one following or simultaneous with the display. Their interpretation is in line with the present one, namely that the cue affects the order of transfer of items from iconic to short-term memory. Their data, the present data, and those of others (e.g., Mewhort, Merikle, & Bryden, 1969) all suggest that transfer is a serial mechanism.

At the same time, however, the results of the masking manipulation in the present experiments provide clear evidence for a parallel mechanism. Furthermore, other data suggest the existence of parallel mechanisms. For example, McIntyre, Fox, and Neale (1970) asked *S*s to detect a target letter embedded in a matrix of noise letters. They varied the redundancy of the noise elements to each other and the similarity of the noise to the target letter. They found that increased redundancy and decreased similarity both served to increase the proportion of correct detections over a control condition. The authors suggest that redundancy has its effect at a point in the system where the input is still parallel. Their interpretation is also consistent with the findings of Dick (1970) who showed that the amount of structural redundancy was not related to postexposure decay but was related to performance.

In summary, the present data provide evidence for both serial and parallel mechanisms. This suggests that the question whether stimuli are processed in serial or parallel bypasses more important issues. A similar point has been made by Garner (1970) in discussing separable and integral dimensions. Inasmuch as the present data provide evidence for both serial and parallel processing on the same set of stimuli, Garner's point may be generalized. In addition to the kind of stimuli employed, one must consider the points in the processing sequence that are being sampled. There is also the intriguing possibility that differences between parallel and serial processing could be used as a tool of analysis of the organization of sensory and cognitive processes.

REFERENCES

DICK, A. O. Visual processing and the use of redundant information in tachistoscopic recognition. *Canadian Journal of Psychology*, 1970, **24**, 133–141.

DICK, A. O. On the problem of selection in short-term visual (iconic) memory. *Canadian Journal of Psychology*, 1971, **25**, 250–263.

ERIKSEN, C. W., & COLEGATE, R. L. Selective attention and serial processing in briefly presented visual displays. *Perception & Psychophysics*, 1971, **10**, 321–326.

ERIKSEN, C. W., & ROHRBAUGH, J. W. Some factors determining efficiency of selective attention. *American Journal of Psychology*, 1970, **83**, 330–342.

GARNER, W. R. The stimulus in information processing. *American Psychologist*, 1970, **25**, 350–358.

HARCUM, E. R. Interactive effects within visual patterns on the discriminability of individual elements. *Journal of Experimental Psychology*, 1964, **68**, 351–356.

HOLDING, D. H. The amount seen in brief exposures. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 72–81.

MCINTYRE, C., FOX, R., & NEALE, J. Effects of noise similarity and redundancy on information processed from brief visual displays. *Perception & Psychophysics*, 1970, **7**, 328–332.

MEWHORT, D. J. K. Familiarity of letter sequences, response uncertainty, and the tachistoscopic recognition experiment. *Canadian Journal of Psychology*, 1967, **21**, 309–321.

MEWHORT, D. J. K., MERIKLE, P. M., & BRYDEN, M. P. On the transfer from iconic to short-term memory. *Journal of Experimental Psychology*, 1969, **81**, 89–94.

NIESSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

SPERLING, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, **74** (11, Whole No. 498).

TOWNSEND, J. T., TAYLOR, S. G., & BROWN, D. R. Lateral masking for letters with unlimited viewing time. *Perception & Psychophysics*, 1971, **10**, 375–378.

WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

(Received January 17, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 67–72

FRUSTRATION AND TASK COMPLEXITY: AN EXTENSION OF FRUSTRATION THEORY [1]

J. WESLEY LIBB [2]

University of Alabama

Performance on a complex task immediately following frustrative nonreward was investigated with 45 college students and findings partially replicated. After an expectancy for reward was established in a two-task sequence, the first task was reinforced intermittently and number of errors and latencies on the second more complex task analyzed. During the early stage of learning more errors occurred on the second task following nonreward than following reward on the first task. In later stages of learning, there was a tendency for fewer errors to follow nonreward than reward, but this finding was not replicated. In the overall within-*S* analyses, a greater number of errors followed frustration. Such differences were not maintained in between-*S* analyses, suggesting that frustration may affect performance but not learning.

Amsel's (1958) frustrative nonreward theory postulates that nonreward for a response previously rewarded elicits a primary frustration reaction (R_f). Feedback from this reaction leads to enhanced motivational effects evident in subsequent instrumental behavior. The alleged motivational effects are typically seen in the invigoration of responses following the nonreward experience. This increment in performance has been termed the "frustration effect" (FE). Numerous studies attest to the empirical validity of this aspect of frustration theory with both animals and children (Ryan & Watson, 1968; Wagner, 1959). Data for adult human *S*s are lacking, perhaps due to difficulties encountered in providing material incentives that outweigh *S*s' self-induced symbolic motivations.

Frustration effects have generally been demonstrated in a task with few obvious competing responses and with a response well established by previous training. On more complex tasks with more obvious competing responses, effects may differ. Among human *S*s, the effects of various levels of drive on the performance of complex tasks has been extensively investigated by Spence (1958). For Spence, working within the Hull-Spence system, response strength is assumed to be a function of excitatory potential (E), which is a function of the interaction of generalized motivational construct (D), an incentive construct (K), and a hypothetical learning construct (H). Additionally, inhibitory factors present in any situation introduce variability in E values producing a normal probability distribution of momentary E values. In order for a response to occur, the momentary excitatory potential of that response must exceed a minimum or threshold (L) value. The probability that a response will occur on any particular trial is a function of the proportion of its normal distribution of momentary E values which extend above L. On a simple or noncompetitive task, increases in D resulting from frustration will raise the E probability value for the correct response farther above L. Although E probabilities for competing responses will also be raised above L, this detrimental effect will be offset largely by the magnitude of differences of E values for correct versus incorrect responses. In a complex task, there are more competing responses present and E values are closer to the magnitude of the correct response, leading to the increased probability that a competing response will occur instead of the correct response on any particular trial. Thus within the Hull-Spence system, drive-enhancing events immediately preceding the performance of a task should lead to enhanced performance if the task is simple and to impaired performance if the task is complex.

---

[1] The author wishes to express thanks to Camella Serum and Thomas Amster for assitance in this experiment.

[2] Requests for reprints should be sent to J. Wesley Libb, Ridgecrest Children's Center, Box 2232, University, Alabama 35486.

Task performance following drive-enhancing events may change over time in the experimental procedure as the developing habit strength of the correct response moves from near equality with competing responses to dominance. Frustration in the early stages of learning may impair performance; frustration in the later stages may become beneficial. When the task to be learned is itself simple, the effects of frustration would be detrimental for only a very brief period. On a complex task where there are many competing responses of nearly equal habit strength, the motivational increase resulting from frustration might not enhance performance until very late in training, if at all.

Schmeck and Bruning (1968) lend support to the hypothesis that frustration produces impaired performance on a complex task. The relationship between frustration and performance in their study is not a simple nor clear one. In fact, their finding a greater number of errors following frustration than following reward after many learning trials seems to run counter to predictions from the Spence model. The task should grow easier as learning progresses; hence, frustration effects should be less obvious or perhaps evidenced in fewer errors.

The experiments that follow were designed to investigate performance on a complex task as it varies with training. High intratask competition is promoted by introducing a learning task with four highly similar responses from which to choose the correct response depending upon which of four similar cues are presented. Specifically, the following hypotheses are investigated: (*a*) Early in learning a complex task, when the difficulty level of the task is high, human *S*s will make more errors following nonreward than following reward; (*b*) in later stages of learning, when the associative strength of the correct response has been established, frustration will be less detrimental and may be associated with fewer errors, that is, an interaction over trial blocks is predicted. Predictions of latency scores following frustrative nonreward early in learning a complex task is without foundation in theory or previous research. Reduced latencies would be anticipated when the task is well learned.

Difficulties in ascertaining the precise point at which a complex task becomes a "simple" task suggest problems in determining where in the data to look for the supposed interaction effect. Pilot work indicated that most *S*s in the experimental task employed in these experiments learned the task during the first block of 10 trials; hence, the second block of 10 trials was assumed to represent the point at which the task became simple for all *S*s.

EXPERIMENT I

Method

Subjects.—Forty-five introductory psychology college students from the University of Alabama volunteered to serve as *S*s to gain extra course credit.

Apparatus.—The experimental chamber was a sound dampened 10 × 10 ft. (3 × 3 m.) cubicle. Electromechanical programimng equipment was located in a separate room. Manipulanda and the reward dispenser were mounted in the experimental chamber. Directly above a small table a Gerbrands Universal Feeder was mounted from which poker chips were dispensed for appropriate responses in a two-task sequence. The first manipulandum in position on the table was a single button (1 in. or 2.54 cm. in diameter) mounted on an 8 × 10 in. (2 × 2.5 dm.) panel. Immediately above the button on the panel a small 28-w. stimulus light was mounted. The second panel, 8 × 14 in. (2 × 3.5 dm.) in size, separately mounted to the right of the first, consisted of four buttons, centrally positioned below a multiple-stimulus projector (Grason-Stadler, #4580-159), also mounted on the panel. Complexity on the second task, as part of a two-task sequence paralleling Amsel's (1958) apparatus, was introduced by varying the stimuli presented on the multiple-stimulus projector which cued *S* as to the correct button on that trial.

Procedure.—The *S*s participated to earn a predetermined number of poker chips to qualify for extra credit in course work. They were divided at random into three groups. All experienced 5 training trials and 20 subsequent test trials in an experimental design similar to that of Schmeck and Bruning (1968). A trial consisted of a sequence of re-

TABLE 1

Mean Errors and Latencies Per Trial Block

Group	Error			Latency		
	Block 1	Block 2	Total	Block 1	Block 2	Total
100–50R	2.867	1.467	4.33	1.563	1.764	1.663
100–50NR	5.067	.867	5.93	1.668	1.317	1.493
100–100	4.000	1.267	5.27	1.307	1.248	1.223
0–0	4.267	1.833	6.10	1.280	1.167	1.277

sponses on both panels. The sequence started with illumination of the stimulus on the first panel. Ten seconds after *S* responded on this panel, the multiple-stimulus projector on the second panel was illuminated permitting *S* to choose one of four buttons. The sequence recycled 30 sec. after the appropriate response on the second panel. Reinforcement consisted of the delivery of one poker chip on rewarded trials. On training trials for the second task, only one stimulus was presented, and the first button was always the correct response. Following training trials, the stimuli on the projector were scrambled and *S* required to learn the appropriate button to be paired with red, blue, yellow, or green lights. On this panel, *S* could press as many buttons as he wished, as many times as he wished, but reward was forthcoming only when the correct button was pressed. Poker-chip reinforcement was continuous on the second panel for all *S*s. On the first panel the experimental group (Group 100–50) received reinforcement 100% of the time during training and 50% randomized reinforcement during test trials. Reinforcement was randomized for each *S* so that 5 such trials occurred in the first block of 10 test trials and 5, in the last block of 10 trials. No more than 3 trials of the same kind in sequence were permitted to occur. One control group (100-100) received 100% reinforcement during both training and test conditions while another control group (0–0) received reward on neither training nor test trials, but were required to respond on the first task to introduce the second task where reward was continuous. The *S*s were given 15 min. to earn a variable number of poker chips so as to qualify for course credit. The training procedure was described, and *S*s were told they would be required to learn the rewarded button on the second panel after a brief pause following the five training trials.

Errors and latency to the first response on the second task were recorded for each trial. Response latency was defined as the time from illumination of the multiple-stimulus projector to *S*'s first button press, either correct or incorrect.

Results

The mean errors for Group 100–50 were combined into two blocks of five trials each (Table 1), as previously discussed. The first statistical analysis was computed to determine performance differences of the experimental group following reward and nonreward over the two blocks of trials. This was computed by a Treatment × Treatment × *S*s analysis of variance which indicated that number of errors following nonreward was greater than following reward, $F\ (1,\ 14) = 5.40$, $p < .05$. The trial blocks effect was also significant, $F\ (1,\ 14) = 35.1$, $p < .001$, indicating a highly significant difference in performance in the first and last blocks of trials.

The Trial Blocks × Reinforcement interaction was also significant, $F\ (1,\ 14) = 11.40$, $p < .01$. The *S*s made significantly more errors following nonreward trials than following reward trials in the first block of trials, $t(14) = 4.77$, $p < .01$. The unusually low error score following rewarded trials in this block of trials was not anticipated and raises questions regarding the representativeness of these data. Though a tendency for fewer errors following nonreward as opposed to errors following reward on the last block of trials was observed, this difference was not statistically significant, $t(14) = 1.25$.

A between-*S* analysis of variance, comparing the mean errors of Group 0–0 with those of Group 100–50 on nonrewarded trials only, revealed no main effect difference. (On all between-*S* comparisons, the error score per trial block used in the analysis of variance was one-half of the total errors for that trial block. Thus number of trials was equated for experimental and control *S*s.) The within-*S* trials block effect was significant, $F\ (1,\ 28) = 94.93$, $p < .001$, and the Trial Blocks × Groups interaction was significant, $F\ (1,\ 28) = 6.73$, $p < .01$. In the first block of nonreward trials, Group 100–50 made significantly more errors than Group 0–0, $t(14) = 2.35$, $p < .05$, and made significantly fewer errors in the final block of trials, $t(14) = 2.84$, $p < .01$.

The between-*S* analysis of variance comparing the mean errors of the 100–100 group with those of the 100–50 group on rewarded trials only revealed no main effect difference while the within-*S* trial blocks effect was again significant, F (1, 28) = 24.24, $p < .001$. The F ratio for the interaction between trial blocks and groups was not significant at an acceptable level, F (1, 28) = 2.52.

The latency data were analyzed in a fashion similar to the error scores. Only one analysis proved statistically significant: the trial blocks main effect with Group 100–50 rewarded trials and Group 100–100 where latencies decreased significantly over trials, F (1, 28) = 10.19, $p < .01$. Within-S data were generally consistent with previous frustration literature in that reduced latencies followed frustrated trials, though this effect was not statistically significant. This effect was actually reversed in between-S comparisons. Perhaps the 10-sec. latency was not optimal for demonstration of the effect. Data from frustration studies with children indicate the optimal time to be between 5 and 10 sec. (Ryan & Watson, 1968). Since latency effects in the frustration literature were observed when the second task was a relatively simple task, the failure to obtain a latency effect may be related to the complex task procedure.

EXPERIMENT II: A PARTIAL REPLICATION

Support for Hypothesis (*a*) in Exp. I may have arisen primarily from the unusually low error scores obtained following rewarded trials in Block 1 for the 100–50 group, raising questions of the validity of the data. The data from this group might be questioned in light of difference from the Schmeck and Bruning (1968) data with infrahuman *S*s. Additionally the nonsignificant latency data suggested a reduction of the intertask interval. A replication of the 100–50 group phase of this experiment was in order. Additional control groups were considered unnecessary inasmuch as differential predictions were related primarily to within S variables.

Method

Subjects and apparatus.—Eighteen introductory psychology college students from the University of Alabama volunteered to serve as *S*s to gain extra course credit. The apparatus was identical to Exp. I, except that marbles were dispensed instead of poker chips.

Procedure.—The *S*s participated to earn a predetermined number of marbles to qualify for credit in course work. As in Exp. I, they experienced five training trials and were then introduced to test trials with 50% nonreward on Task 2, randomized within three trial blocks of eight trials each. No more than two trials of a kind occurred in sequence, and the initial trial in the test phase, i.e., reward or nonreward, was balanced across *S*s. The intertrial interval was reduced from 10 sec. to 6 sec. since the literature suggests that a reduced interval might be a more viable procedure for evidencing a latency effect. Instructions were modified to accommodate minor procedural changes. In all other respects, the procedure was identical to Exp. I.

Results

The mean errors per trial block for all *S*s were combined into three blocks of four trials each (four reward and four nonreward for each block of eight trials) and are presented in Table 2. Data were analyzed in a Treatment × Treatment × S analysis of variance. Although a significant trial blocks effect was obtained, F (2, 17) = 54.5, $p < .001$, the overall error scores following nonrewarded trials did not differ significantly from errors following rewarded trials, nor was the interaction with trial blocks significant, F (2, 17) = 3.15, $p < .065$. Since specific predictions had been made about direction and location of effects within the experimental design, a t-test analysis of scores in the first block of trials was performed. In this analysis, significantly more errors followed nonrewarded than rewarded trials in Block 1 $t(17) = 1.82$, $p < .05$. No significant differences emerged in the latency data.

DISCUSSION

Hypothesis (*a*) was supported in Exp. I in both within- and between-S analyses.

TABLE 2

MEAN ERRORS PER TRIAL BLOCK

Group	Block 1	Block 2	Block 3	Total
100–50R	3.50	1.1	.33	4.94
100–50NR	4.72	1.1	.61	6.44

Where nonreward occurred early in learning prior to performance on a task with high intratask competition, *S*s made more errors than when this task followed a rewarded task. The within-*S* finding was replicated in Exp. II. Support for Hypothesis *b* is more tentative. A significant interaction in Exp. I was obtained in a between-*S* analysis of data, and *S*s in the 100–50 group not only made significantly more errors early but significantly fewer errors later in learning a complex task on trials following nonreward than did *S*s rewarded continuously on the prior task. Although the within-*S* effect was consistent with the between-*S* difference, this effect was not replicated in Exp. II where error scores in both Blocks 2 and 3 were nearly identical for rewarded and nonrewarded trials. Closer inspection of the data from the second experiment revealed that one *S* made as many errors in the third block as in the first block of trials, suggesting random responding throughout all 24 trials. Removing scores for this one *S* yielded data more consistent with Exp. I. Perhaps in a multiple-choice situation the task did not become a "simple task" comparable to that for a rat traversing a runway. This generalization may be supported by the failure of the latency data to approximate Amsel's (1958) findings. In Exp. I, *S*s made more errors overall when frustrated than when not frustrated; however, this finding was not replicated at an acceptable level of confidence in Exp. II. However, data trends were consistent. Significance in Exp. I may have been due to a spuriously low error score following rewarded trials in Block 1 for the 100–50 group.

A major difficulty encountered in this experiment is in the identification of the point at which the task becomes a "simple" or noncompetitive task. Since *S*s learn at different rates, simply lumping scores into blocks of trials and arbitrarily identifying one set of scores as the point at which the task becomes simple may be an insensitive procedure. It was felt that the second and third blocks of trials in Exp. II would present situations in which several degrees of learning were present. Removing the data for *S* who responded randomly revealed that only five errors were made by the remaining 17 *S*s in the third block of trials. In the second block of trials, errors occurred with greater frequency: 17 following rewarded trials and 11 following nonrewarded trials. Perhaps an even earlier point of analysis would have yielded a significant reversal of the error data; however, *S* differences in learning would not permit this since the "early" learning frustration effect likely would have entered as a confounding factor. The data were reviewed utilizing a backward-learning curve; however, this treatment of the data revealed nothing of interest, and the fact that errors sometimes appeared after several errorless trials suggests this procedure to be inappropriate to the data.

In general the results of these experiments are seen as consistent with the Hull-Spence system and extend Amsel's (1958) frustration nonreward theory to include the complex task situation. The fact that a greater number of errors followed nonreward in the within-*S*, but not in the between-*S* analysis of error scores may suggest that the observed within-*S* differences are performance differences only and that learning proceeds at the same rate regardless of preceding frustration, consistent with previous research on the effects of enhanced drive.

The data are generally consistent with Schmeck and Bruning (1968). The major difference in findings is the error score interaction, with fewer errors occurring after the task is relatively well learned. Such differences may be related to *S* differences, or task differences or the possibility that *S*s in Schmeck and Bruning had not learned the task to the degree required to produce this interaction. Since validity of the interaction effect is somewhat tenuous, further clarification is in order.

Demonstration of frustration effects with adult human *S*s is unique and may have been a function of the procedure employed to motivate *S*s. Failure to demonstrate such effects in latency data may suggest that this measure is insensitive to enhanced drive among adult human *S*s when performing on a complex task. Human *S*s come to the experimental situation with a greater complexity of preexperimental history, frequently labeled as personality variables. Differentiation of *S*s on the basis of these personality variables may permit refinement of experimental procedures and clarification of within-*S* consistency in response to frustration.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

RYAN, T. J., & WATSON, P. Frustrative nonreward theory applied to children's behavior. *Psychological Bulletin*, 1968, **69**, 111–125.

SCHMECK, R. R., & BRUNING, J. L. Task difficulty

were much more practiced by Nickerson's (1965) *S*s than were those comparison operations resulting in a DIFFERENT response.

This experiment was performed in order to investigate the effects of varying amounts of practice with particular pairs of stimuli on SAME–DIFFERENT RTs. It was anticipated that differences in frequency of exposure to particular stimulus pairs would strongly affect RTs because of differences in the amounts of practice with the unique comparison operations associated with each stimulus pair. Therefore, "same" RTs might be made faster or slower than DIFFERENT RTs simply by varying the amount of practice with the appropriate stimulus pairs.

METHOD

Subjects.—The *S*s were 60 Ohio State University undergraduates fulfilling a course requirement. Each *S* was assigned randomly to one of four experimental groups using a table of random numbers. Each *S* served in the experiment for a single half-hour session.

Apparatus.—Two Kodak Carousel RA–950 random-access projectors simultaneously displayed pairs of stimulus slides. The slides were displayed simultaneously because of equipment restrictions during the time of the experiment. It has been demonstrated that the results obtained by simultaneous presentation are generally quite similar to those resulting from sequential presentation (Nickerson, 1967). The stimulus pairs were displayed for 2.5 sec. A photoelectric diode activated by the opening of the projector shutter triggered a Hewlett-Packard 52236 electronic timer, and *S*'s voice keyed a Schmitt trigger which stopped the timer. The *S*s received white noise over earphones to mask extraneous noises, and *E* could hear all of *S*'s responses.

The slides were back projected on a ground-glass screen, 2 ft. before *S*. The two stimulus slides together measured 9 × 4 in. with a 1-in. space between the two slides.

Stimuli.—The stimuli were 15 eight-sided forms selected from Vanderplas, Sanderson, and Vanderplas (1965) and had association values between 20% and 45%. Each *S* worked with five of the forms during the experiment, and the stimuli used were counterbalanced across *S*s within each group. RTs were recorded for both SAME and DIFFERENT responses, and two data analyses were performed: (*a*) only the RTs which resulted from the comparisons of two preselected stimuli for each *S* were analyzed in the first analysis, (*b*) the combined data from all stimulus pairs were analyzed in the second analysis. In this manner, *S*s served as their own controls since they provided RTs for the analysis of the preselected stimuli as well as for the analysis of all stimuli combined. The actual preselected stimulus pairs for which RTs were analyzed were counterbalanced across *S*s such that each stimulus was measured twice in each group.

Procedure.—In order to control the amount of exposure to the preselected stimulus pairs, four experimental groups were formed. The *S*s in Group S5D viewed the preselected stimulus pairs which resulted in a SAME judgment five times more often than they viewed the preselected nonmatching stimulus pairs. For this group, Stimulus Pairs A-A and B-B were each displayed 40 times, and Pairs A-B and B-A were each displayed 4 times. (In order to simplify the explanation of the experimental procedure, the two preselected random-form stimuli for an *S* will be called Stimuli A and B, and the first stimulus noted in a pair is the stimulus which appeared on the left of *S*'s display.) Since the Pair A-B should be different along the same relationships as Pair B-A, it was assumed that the unique comparison operation for Stimulus A with B was performed 8 times to the 40 times for Stimulus Pairs A with A and B with B. It was anticipated that SAME RTs for the preselected stimulus pairs would be *faster* than DIFFERENT RTs for Group S5D because the two preselected stimulus pairs resulting in a SAME judgment were presented five times more often than were the preselected pairs resulting in a DIFFERENT response. This procedure also assured that each of the individual preselected stimuli was utilized 48 times. In order to balance the probability of SAME and DIFFERENT responses for each *S*, 40 each of the matching pairs of the remaining three stimuli and 192 nonmatching pairs of the remaining three stimuli were also viewed by *S*, yielding 200 SAME and 200 DIFFERENT responses during the session. The treatment conditions for this group and the other three groups are outlined in Table 1.

Group S1D viewed the preselected stimulus pairs resulting in a SAME judgment an equal number of times as the DIFFERENT pairs. Stimulus A-A and B-B were presented 24 times each, and Pairs A-B and B-A were viewed 12 times each. Again, the remaining three stimuli for the group were used to balance the probability of the responses, resulting

TABLE 1

FREQUENCY OF PRESENTATION OF PRESELECTED STIMULUS PAIRS FOR EACH GROUP

Group	Frequency of "same" stimulus pairs		Frequency of "different" stimulus pairs		No. of trials each stimulus used	
	A-A	B-B	A-B	B-A	A	B
S5D	40	40	4	4	48	48
S1D	24	24	12	12	48	48
D5S	8	8	20	20	48	48
D10S	4	4	20	20	44	44

in 120 SAME and 120 DIFFERENT responses for a session.

Group D5S viewed the preselected Pairs A-A and B-B 8 times each, while Pairs A-B and B-A were each presented 20 times. Therefore, for this group, it was assumed that the unique comparison operation for Pairs A-B and B-A was used five times more often than the comparison operation for pairs A-A and B-B. Pairs were added from the remaining stimulus forms to make a total of 52 SAME and 52 DIFFERENT responses.

The *S*s in Group D10S viewed the Pairs A-A and B-B 4 times each and the preselected pairs A-B and B-A 20 times each. Therefore, it was assumed that the unique comparison operation for Pairs A-B and B-A was used 10 times more often than the comparison operation for Pairs A-A and B-B.

RESULTS

An analysis of variance using the median RT data from all stimulus pairs (the control analysis) revealed no significant effect of the four experimental groups, F (3, 56) = 1.40, or of the Groups × Response Type interaction, F (3, 56) = .500. However, a significant main effect of response type was found, F (1, 56) = 5.638, $p <$ .025. The data making up this analysis are displayed in Fig. 1. This figure shows that, as in Nickerson's (1965) results, the median RTs for SAME stimulus pairs were shorter than those for different pairs.

A second analysis of variance was performed using the median RT data from the preselected stimulus pairs. This analysis revealed no significant effect of groups, F (3, 56) = 1.22, or of response type, F (1, 56) = 1.67. However, a significant Groups × Response Type interaction was found, F (3, 56) = 10.612, $p < .001$. The data making up this interaction are displayed in Fig. 2. This figure shows that different amounts of practice with specific preselected stimulus pairs produced differences in SAME/DIFFERENT RTs for those pairs. When the preselected pairs resulting in the judgment of sameness were presented five times more often than the "different" pairs (Group S5D), the SAME RTs were faster than the DIFFERENT RTs. The reverse was true when the pairs resulting in the response DIFFERENT were presented more often (Group D10S). These results cannot be attributed to the familiarity of a particular response because both positive and negative responses were equally probable for each group. Analysis of the simple effects of responses between preselected SAME and DIFFERENT RTs within each group indicated that there were significant differences between these responses for Groups S5D ($p < .01$), S1D ($p < .05$), and D10S ($p < .01$).

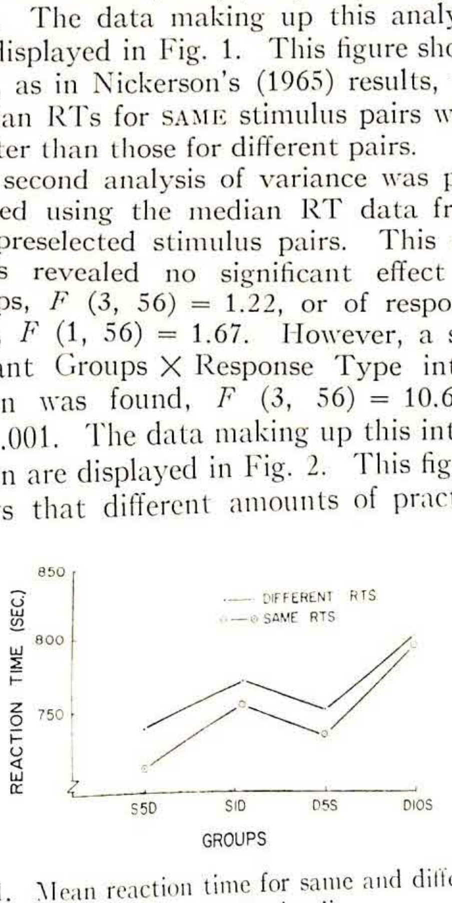


FIG. 1. Mean reaction time for same and different responses—all stimuli.

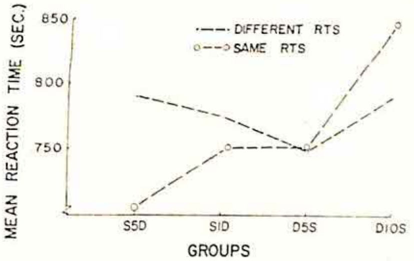


FIG. 2. Mean reaction times for same and different responses—preselected stimuli only.

The data generally support the hypothesis that the shorter RTs for SAME responses found by Nickerson (1965) and found in this experiment when the data from all RTs are analyzed were caused by a greater amount of practice with the stimulus pairs resulting in a judgment of sameness. However, it is surprising that there were no significant differences between the two types of responses in Group D5S, which actually experienced a more extreme discrepancy in the number of times the preselected SAME and DIFFERENT stimulus pairs were viewed than did the S1D group. Also, it is surprising that the differences between response types in Group D10S were less than the differences found for Group S5D. The mean difference in RTs for Group D10S (53 msec.) was 33 msec. less than the mean difference for Group S5D. Why should the difference between SAME and DIFFERENT RTs be less for Group D10S than for Group S5D, when D10S experienced the more extreme differences in exposure to the preselected pairs? Also, why should the differences be greater for group S1D than for D5S?

In an attempt to answer these questions, a critical look was given to an assumption of the experiment. It was assumed that the unique comparison operation used for the Stimulus Pair A-B was the same operation used for Pair B-A. Both left–right orders were assumed to be mutually cumulative in their contribution to practice effects for the comparison operation. However, if this were an incorrect assumption, it follows that the comparison operation for Pair A-B would be different from that used for Pair B-A. Each left–right order would evoke its own unique comparison operation and each operation would contribute only to its own backlog of practice. Therefore, the amount of practice with stimulus pairs *in their particular right–left order* would be the important variable, and the differences in amounts of practice with the matching and nonmatching preselected stimulus pairs would not be those calculated for the groups at the beginning of the experiment.

The revised ratios of the amounts of practice with matching and nonmatching preselected stimulus pairs are evident from inspection of Table 1. Under the new assumption, Group S5D actually performed the SAME unique comparison operations 10 times more often than the DIFFERENT operations. Group S1D practiced the SAME operation twice as often as the DIFFERENT operation, Group D5S practiced the different operations only twice as often as the matching operations instead of five times as often; and Group D10S only practiced the nonmatching operations to specific preselected pairs five times as often as the matching operations to preselected pairs. For example, in Group S5D, the Stimulus Pairs A-A and B-B were each presented 40 times, while Pairs A-B and B-A were each presented 4 times, producing a revised ratio of practice with measured matching pairs to nonmatching pairs of 10:1.

These new practice ratios explain half of the questions raised above. The difference between SAME and DIFFERENT RTs was greater for Group S5D than for Group D10S because *S*s in Group S5D actually experienced a more extreme difference in practice with the different comparison operations. However, since Groups S1D and D5S would have actually experienced similar practice differentials for the comparison operations, the mean absolute differences between the two types of responses should be the same for both groups. Figure 2 illustrates that this is clearly not the case.

DISCUSSION

The results of this study demonstrate that the relationship between SAME and DIFFERENT RTs is very much influenced by the amount of experience which *S* has had with the specific pairs of stimuli to which he is responding. The present author's contention is that each stimulus pair used for comparisons generates a unique comparison operation and that practice effects accumulate for each unique operation independent of other comparisons which may take place. It should be noted that the practice effects may be specific to the right–left arrangement of the stimulus pairs. This suggests that each *S* must perform his visual search in an invarying right–left or left–right order, and that the order of comparison of the two stimuli affects the unique comparison operation which is performed. This is not an unreasonable suggestion considering the results of Sperling (1960).

The interpretation of a unique comparison operation is preferred over the possibility that *S*s may be responding to the entire display (stimulus pair) as a gestalt instead of performing a comparison between the two stimuli. The high probability of a comparison operation being performed was demonstrated by Nickerson (1967), who found that sequential or simultaneous presentation of the stimulus pairs did not alter the relationships between SAME and DIFFERENT RTs.

The notion of practice with unique comparisons of stimulus pairs also explains another result reported by Nickerson (1967). In his study, he used geometric figures varying with respect to one, two, or three dimensions. He found that RTs increased when the number of dimensions upon which his stimuli differed was reduced from three to one, but that SAME responses (generated when the stimuli differed on zero dimensions) were shorter than would be expected from an extrapolation of the nonmatching stimulus data. Since Nickerson (1967) used all combinations of 27 stimuli, balancing the number of match and nonmatch cases, *S*s' practice with specific matching stimulus pairs was much more extensive than that with specific nonmatching pairs. Therefore, the SAME RTs would be depressed because of the greater amount of practice with the unique comparison operations associated with matching stimulus pairs.

REFERENCES

NICKERSON, R. S. Response times for same–different judgments. *Perceptual and Motor Skills*, 1965, **20**, 15–18.

NICKERSON, R. S. Same–different response times with multi-attribute stimulus differences. *Perceptual and Motor Skills*, 1967, **24**, 543–554.

SPERLING, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, **74**, (11, Whole No. 498).

VANDERPLAS, J. M., SANDERSON, W. A., & VANDERPLAS, J. N. Statistical and associational characteristics of 1100 random shapes. *Perceptual and Motor Skills*, 1965, **21**, 414.

(Received January 26, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 78–86

WORDS AND PICTURES IN AN STM TASK[1]

WILLI TERNES AND JOHN C. YUILLE[2]

University of British Columbia, Vancouver, British Columbia, Canada

Verbal and visual short-term coding processes were investigated by visually presenting word or picture stimuli and testing recall and recognition performance at zero delay as well as after rehearsal and verbal interference intervals. Recall and recognition results for verbal material replicated previous studies investigating verbal short-term memory (STM). Similarly, picture recall data reflected verbal STM characteristics. In contrast, picture recognition results could not be incorporated within a verbal STM framework, and a visual coding system is postulated to account for the data.

The use of verbal material in short-term memory (STM) research has led theorists to consider STM as a predominantly acoustic or verbal–auditory system (Atkinson & Shiffrin, 1968; Neisser, 1967). Characteristic features of the verbal STM include a limited storage duration and capacity, which is assumed to be a function of the verbal rehearsal mechanism, and a proneness to interference during short-term retention intervals filled with interpolated activities such as auditory masking or counting backwards (Atkinson & Schiffrin, 1968). Recently, evidence suggesting the existence of a similar visual STM has been obtained. For example, physical matching was shown to be more rapid than name matching for visually presented letters (Posner, Boise, Eichelman & Taylor, 1969). Tversky (1969), using face drawings and nonsense names, demonstrated that verbal and visual coding strategies can be manipulated by *S*s' expectations of the test modality. However, these studies tested visual STM immediately following stimulus presentation, and evidence for visual coding over short-term retention intervals appears less consistent. Parks, Parkinson, and Kroll (1971), for instance, show that retention of visual information is found after an auditory masking task. In addition, Cohen and Grandstrom (1970) report that recognition of pictorial material is lower following a visual interpolated activity than after rehearsal or verbal interpolated activities. On the other hand, there is some evidence that performance improves over unfilled retention intervals. Milner (1968) obtained an increase in recognition of faces, and Posner (1967) found an increase in accuracy with a visual location response with rest intervals. In contrast, Potter and Williams (1970) have introduced evidence against a rehearsal mechanism in visual STM.

To summarize, while there has been evidence supporting the existence of a visual STM, characteristic features of this visual code over retention intervals filled with various types of interpolated activities are not yet consistently demonstrated. One major problem inherent in the investigation of a visual as distinct from a verbal STM is that of dual coding, that is, the extent to which visual material is also coded verbally and, conversely, verbal material is coded visually. Apart from the specific interpolated activity used in a given experiment, there can be little argument that the nature of the visual stimulus, the duration of the stimulus presentation, as well as the duration of the retention interval, provide critical sources for dual coding of verbal as well as visual stimuli. In effect, it becomes necessary to distinguish between recognition of visual information in contrast to recognition based on visual short-term coding. For instance, with relatively long stimulus displays, visual recognition might be a factor of both visual and verbal coding processes, while recognition with shorter visual displays might be based predominantly on visual coding. Paivio and Csapo (1969) have demonstrated convincingly that, at least with familiar objects in their verbal and pictorial form, a substantial reduction in dual coding cannot be achieved until the stimulus presentation is shortened to around 200 msec. With respect to the aboved mentioned experiments, only the Potter and Williams (1970) study approaches this experimental restriction. It is not surprising, therefore, to note the different conclusions about the visual rehearsal effect reached by Potter and Williams, in contrast to Milner (1968) and Posner (1967).

---

[1] This research was supported by grants to the second author from the National Research Council of Canada (#327) and the University of British Columbia. The paper is based on a thesis submitted by the first author to the University of British Columbia in partial fulfillment of the master's degree.

[2] Requests for reprints should be sent to John C. Yuille, Department of Psychology, University of British Columbia, Vancouver 8, Canada.

It is the purpose of the present research to explore further ways of contrasting verbal and visual, or at least predominantly verbal and visual, coding systems in STM. In light of the above mentioned arguments, a stimulus situation was created to meet the following two criteria: (*a*) a reduction in dual coding processing had to be achieved; and (*b*) a zero-delay response level had to be obtained which would allow for possible increases or decreases due to experimental manipulations. A series of pilot studies was undertaken utilizing the stimulus material tested by Paivio and Csapo (1968). Stimulus complexity was varied from one item to a matrix of nine items over a series of stimulus durations. A matrix of six items presented for 1.2 sec. was selected as an optimal stimulus situation meeting the above mentioned criteria.

The general question to be answered in this investigation was whether the variables that normally affect performance with verbal material will similarly affect performance with pictorial material. If they would, a single coding system could be postulated for the two types of material; if not, then separate coding systems with qualitatively or quantitatively different characteristics would have to be proposed.

The research was designed to assess short-term coding by investigating recall and recognition performance for visually presented words and line drawings of familiar objects. Performance after zero delay was examined to provide an assessment of relative coding efficiency for verbal and pictorial material. Furthermore, the two coding systems were contrasted under rehearsal and interference conditions over short-term retention intervals. Since rehearsal and interference effects for verbal material are experimentally well founded, the importance of this research centered around the effects of interpolated activities on the retention of pictorial material.

EXPERIMENT I

The first experiment was designed to investigate whether or not evidence for verbal and visual coding systems could be obtained in a recall task employing visually presented words and pictures. Recall was tested either at zero delay or after 15 sec. of interference or rehearsal.

Method

Subjects.—Sixty introductory psychology students (30 males) volunteered to serve as *S*s.

Material.—A basic pool of 81 stimuli was obtained from stimulus materials used by Paivio and Csapo (1969). The items consist of pictorial representations of familiar objects in the form of simple line drawings as well as their verbal labels. Consistency of labeling data had been obtained on these drawings by Paivio and Csapo to ensure unambiguity of the pictorial stimuli. A subset of 54 items was selected, and these stimuli randomly assigned to nine 2×3 stimulus matrices. Three additional sets of nine matrices each were generated by random combinations of the 54 stimuli. The matrices were photographed in their verbal and pictorial forms, and the negatives mounted as 35-mm. slides with white outlines on a dark background. In addition, two practice matrices were constructed using letters.

Apparatus.—The stimuli were presented with a Kodak Carousel 850 projector. Stimulus duration was controlled by an automatic timer.

Procedure.—The *S*s were individually tested and randomly assigned to one of six groups, with the restriction that the male–female ratio was even in each group. The six groups consisted of three word and three picture conditions. Within each presentation mode, recall was tested at either zero delay, 15-sec. delay with verbal interference, or 15-sec. delay with rehearsal. The four sets of stimulus material were employed equally often in each group. Each stimulus matrix was presented for 1.2 sec., with 60 sec. given for recall. The *S*s received standard instructions for oral recall. Zero-delay groups were instructed to recall the items immediately after stimulus presentation. The interference groups were instructed to count backwards by three's from an aurally presented three-digit number, until the recall cue was given, and rehearsal groups were

TABLE 1

MEAN RECALL SCORES OUT OF A MAXIMUM OF 54 ITEMS

Stimulus presentation	Delay		
	0 sec.	15 sec.	
		Interference	Rehearsal
Word	37.60	23.20	36.00
Pictures	37.60	23.00	32.10

encouraged to rehearse the stimulus items during the retention intervals. To familiarize *S*s with the general procedure, the two practice stimuli were presented before the onset of the test trials. Recall responses were scored manually. For the picture group, a liberal scoring method was employed since items like ship, doll, and oven could equally be labeled boat, girl, and stove, respectively.

Results and Discussion

Mean total recall scores for the nine trials are presented in Table 1. A 2 × 3 analysis of variance with 2 presentation modes (words vs. picture) and the 3 delay conditions as factors, revealed no significant differences between word and picture recall. However, there was a significant main effect for delay conditions, $F\ (2,\ 54) = 55.11$, $p < .01$. Post hoc tests by the Newman-Keuls method showed word and picture recall to be constant between zero delay and 15 sec. of rehearsal. However, after 15 sec. of verbal interference recall for words and pictures showed a significant decrease with respect to both zero delay and rehearsal conditions (all ps $< .01$).

Since word–picture recall differences did not emerge at zero delay, the data can be interpreted as either reflecting a single coding system or as reflecting equally efficient verbal and visual systems. On the basis of interpolated activity effects, however, picture recall data cannot be interpreted in terms of visual coding since rehearsal and interference effects in both groups clearly demonstrate well-established characteristics of the verbal STM (Atkinson & Shiffrin, (1968). That is, loss of information during the rehearsal interval is held to a minimum, while recall performance is significantly reduced when rehearsal is prevented by a verbal interference task. Consequently, word and picture recall data can be adequately explained by a verbal STM.

Since response demands in the experiment require verbal labeling of words as well as pictures, it is plausible that evidence for verbal and visual coding might be obscured by factors inherent in the oral recall situation. Thus, the next experiment examines short-term coding effects in a recognition task, where the need for verbal labeling of pictorial stimuli should be minimized.

EXPERIMENT II

The purpose of Exp. II was to investigate recognition of word and picture stimuli at zero delay and after interference intervals. A factorial design using words and pictures in both the learning and recognition phase generated four groups, i.e., Word–Word (WW), Word–Picture (WP), Picture–Word (PW), and Picture–Picture (PP). The *S*s in the WP group, for example, received words in the learning phase and pictures in the recognition phase. The WW and PP groups were expected to give the most accurate assessment of verbal and visual coding effects. The WP and PW groups were introduced to evaluate the extent to which the presentation mode in the test phase influences recognition performance. The delay conditions were included to determine if verbal interference differentially affects verbal and visual short-term storage. Latency-of-response scores were also obtained to increase the sensitivity of the experiment.

Method

Subjects.—Eighty introductory psychology students served as *S*s (40 males).

Material.—The four sets of stimulus matrices were used from Exp. I. The recognition phase consisted of a sequence of six items per trial, with each item presented individually. Three Old items from the stimulus matrix and three New items from the original stimulus pool constituted the six items. Two sets of recognition sequences were individually photographed and mounted as 35-mm. slides. The four sets of stimuli and the two sets of recognition sequences were combined to ensure that every Old item was presented at two different spatial positions in the simulus matrix, and every spatial position in the matrix was tested with two different Old items per delay condition.

Apparatus.—An automatic timer controlled the stimulus presentation on a Carousel 850 projector. Recognition slides were presented manually. A reaction timer was activated with the presentation of each recognition slide and was stopped when *S* pressed a response-choice button.

Procedure.—The *S*s were tested individually and randomly assigned to one of four groups, i.e., WW, WP, PW, or PP, with the restriction that an equal sex ratio was obtained in each group. Within each group, a subset of five *S*s each received one of the four different sets of material. Each *S* was tested on nine trials, three trials each at 0 −, 15 −, and 30-sec. delay. The delay intervals were randomly assigned to the nine trials in blocks of three, and different random assignments for delay intervals were employed for each of the four sets of material. The *S*s were given a standard set of instructions, differing for the four groups with respect to the particular word and picture combination involved. The WW group, for instance, was presented with a matrix of six words for 1.2-sec. duration. Recognition words were presented one at a time, and *S* was instructed to respond as fast as possible to every recognition slide by pressing one of two buttons: Old if the given word was from the previous matrix, and New if the given word had not appeared in the stimulus display. The *S*s were further told that on some trials the recognition sequence would start immediately following the stimulus presentation, while on other trials, the stimulus display would be followed by a three-digit number from which *S* had to count backwards by three's until the first recognition slide was presented. No indication was given to *S* about the number or sequence of the three types of delay trials, nor about the Old-New item ratio in the recognition sequence. Instructions were followed by two practice trials using letters as stimuli.

The following temporal sequence was maintained throughout the experiment. Each test trial started with a 1.2-sec. presentation of the stimulus matrix, followed within 1 sec. by either the first recognition slide or a three-digit number in the case of delay trials. The presentation rate during the recognition sequence was 7.5 sec. per slide, which included 1 sec. for slide change. There was a 20-sec. interval between the end of the recognition sequence and the onset of the next trial.

TABLE 2

MEAN VALUES FOR BOTH RESPONSE MEASURES FOR EXPERIMENT II AS A FUNCTION OF DELAY CONDITIONS

Cond.	Mean correct responses			Mean d′ scores		
	0	15	30	0	15	30
WW	7.75	6.70	7.60	3.97	3.37	3.17
WP	7.50	6.85	6.95	3.17	2.37	2.20
PW	6.60	6.95	7.20	2.64	3.95	2.95
PP	5.90	7.40	7.45	2.42	2.79	2.96

Note.—WW = word-word, WP = word-picture, PW = picture-word, PP = picture-picture. Column heading 0, 15, and 30 indicate the delay in seconds.

Results and Discussion

Reaction times were almost constant across all experimental groups. Consequently, latency scores were excluded from further analysis. Two types of recognition scores were examined: the number of correct responses and d′ scores incorporating a correction for guessing (Kintsch, 1970). Table 2 presents both scores for each condition. A comparison of these two measures indicates different response patterns across groups and delay conditions which suggest higher guessing rates for WP and PW than for WW and PP groups. Guessing also tended to be higher at longer delay intervals. Separate analyses of variance were performed on these two measures. Each involved a $2 \times 2 \times 3$ analysis with factors: presentation mode in the learning phase (words vs. pictures), presentation mode of the recognition phase (word vs. picture), and 3 delay intervals as variables. Both analyses revealed no significant main effects. However, an interaction between the presentation mode in the learning phase and delay conditions was significant, $F\ (2, 152) = 11.99$; $p < .01$ for number of correct scores, and $F\ (2, 152) = 8.42$, $p < .01$ for d′ scores. Since post hoc tests on the two types of scores differed only at the 30-sec. delay condition, the condition with the highest guessing rate, the discussion will be restricted to the d′ scores in this context. Post hoc tests by the Newman-Keuls method produced three significant comparisons, (all ps $< .01$). For zero-delay conditions, recognition for WW and WP groups was superior to PW and PP groups. Performance for WW and WP groups decreased between zero and 15 sec. of verbal interference. Recognition for PW and PP groups in contrast increased between zero and 15 sec. delay. There were no significant differences between the four groups after 15 or 30 sec. of interference.

While the presentation mode of the recognition phase proved irrelevant, the presentation mode of the stimulus matrix emerged as the determining factor in recognition performance. In general, three separate trends can be observed in the data, only one of which is consistently supported by previous evidence. First, recognition of verbal material suffers from verbal interference. This finding has numerous parallels in previous research (e.g., Murdock, 1963) and is consistent with the recall data in Exp. I. Second, recognition for pictorially presented material increases during a verbal interference task. While these data would not be expected on the basis of any current theoretical model, the evidence at least provides a strong argument against verbal coding in PW and PP groups. Third, there was a significant zero delay recognition difference between words and pictures. Again, this finding is contrary to previous results (Posner et al., 1969; Tversky, 1969). However, methodological differences, e.g., stimulus material and presentation time, could account for the differences.

Hence, in comparison to the recall data of Exp. I, it becomes evident that recognition scores cannot be interpreted in terms of a single coding system. Two general aspects of the recognition data provide strong arguments against a single coding interpretation: (*a*) the difference in recognition scores at zero delay between verbal and pictorial stimulus material, and (*b*) the differential interference effect for the two types of stimulus presentations. While the recognition data in WW and WP display the familiar interference effect of the verbal short-term system, a predominantly nonverbal, or visual short-term system has to be postulated to account for the recognition data of PW and PP groups. Tentatively, the visual short-term system seems less efficient in terms of immediate recognition than the verbal system, and seems immune to the damaging effects of verbal interference.

Since two of the observed trends, i.e., zero-delay differences and the increase in PW and PP groups are relatively unsupported by previous evidence, it was thought necessary to further investigate recognition performance under these conditions. In Exp. III, then, the verbal interference period is replaced by a rehearsal period to supply additional data which can be used to evaluate the results of Exp. II.

EXPERIMENT III

Because both the presentation mode in the recognition phase and the 30-sec-delay condition failed to provide additional information about coding differences in Exp. II, rehearsal effects were investigated only for WW and PP groups at 0, 7.5, and 15 sec. of delay. Latency measures were also discontinued in light of the previous experiment.

Despite these changes in the overall design, the zero-delay conditions remain unchanged and should replicate the results from Exp. II. The facilitative effect of the rehearsal conditions should also be obtained as in the recall experiment. Of prime importance, however, is the recognition date for the picture group after rehearsal intervals. A comparison of the recognition level at 15-sec. delay between the interference (Exp. II) and rehearsal conditions should provide some clarification for two unresolved problems: (*a*) an additional assessment for the interference effect in Exp. II should be obtained, and (*b*) some evidence should be available concerning the need for postulating a visual rehearsal mechanism.

Method

Subjects.—Forty subjects from the same *S* pool as in the previous experiments participated in this experiment (20 males).

Material.—Only one set of material from Exp. II was selected for group presentation. The *S*s were also supplied with scoring sheets to record their own responses.

Apparatus.—An automatic timer controlled the stimulus presentation of the projector. Recognition slides were presented manually.

Procedure.—The *S*s were tested in four sessions, in groups of 10 *S*s each. The general procedure of this experiment was similar to Exp. II, that is, *S*s were presented with either word or picture stimuli, and the three delay conditions were introduced as a within-*S*s variable. The two major changes in this experiment were: (*a*) *S*s were encouraged to rehearse the stimulus items during retention intervals; and (*b*) *S*s were instructed to write down their own recognition responses, i.e., "O" for Old items and "N" for New items. Furthermore, *S*s were tested in groups of 10 *S*s each extending over four sessions. The WW and PP conditions were randomly assigned to these four sessions of group testing.

Results and Discussion

The data were analyzed with respect to recognition responses and d′ scores. A high agreement between the two response measures indicated that a reduction in guessing rates had been achieved by the change in the design. Table 3 shows the response patterns for correct recognition scores and d′ scores. The analyses of variance for both response measures showed significant main effects for presentation mode as well as delay conditions. Again, only d′ scores will be discussed here in detail. Recognition in WW was significantly superior to recognition in PP, F (1, 38) = 16.67, $p < .01$, and performance increased for both groups over delay periods, F (2, 76) = 5.03, $p < .01$.

In effect, the data confirmed that zero-delay responses for WW and PP present reliable recognition differences. These recognition differences at zero delay are maintained over rehearsal intervals by an equally facilitative effect for both presentation modes. Similar rehearsal effects for visual material have already been suggested by Posner (1967) and Milner (1968). However, this initially appealing interpretation of the main effect for delay in terms of a rehearsal effect appears less tenable in conjunction with the data from the previous experiments. For instance, why should word recognition be more efficient after a rehearsal period than immediately following the stimulus presentation in a recognition task, but not in a recall task? Also, why should picture recognition after 15-sec. retention interval show an increase during rehearsal as well as verbal interference (Exp. III vs. II)?

In an attempt to further clarify the nature of these results, as they might be affected by the given task demands of the recognition task, the percentage of correct responses to Old items in the six sequential positions for each recognition sequence was examined for Exp. II and III. It is important to note, though, that results from this response measure must be interpreted cautiously in light of the following two limitations. Percentage data are based on either one or two data points for each of the six positions within each delay condition, and are based on raw data without the benefit of a correction factor.

TABLE 3

Mean Values for Both Response Measures for Experiment III as a Function of Delay Conditions

Cond.	Mean correct responses			Mean d′ scores		
	0	7.5	15	0	7.5	15
WW	7.55	7.25	7.05	3.73	3.51	4.33
PP	5.30	6.35	6.90	2.45	2.75	3.05

Note.—WW = word-word, PP = picture-picture. Column heading 0, 7.5, and 15 indicate the delay in seconds.

TABLE 4

Percent Correct Responses to "Old" Items for Zero-Delay Condition in Recognition Experiments

Exp.	Position					
	1	2	3	4	5	6
II						
WW	90.00	82.50	87.50	65.00	87.50	75.00
WP	100.00	95.00	85.00	45.00	82.50	87.50
PW	80.00	70.00	82.00	65.00	82.50	57.50
PP	85.00	70.00	62.50	55.00	57.50	67.50
III						
WW	95.00	85.00	85.00	80.00	80.00	80.00
PP	90.00	67.50	50.00	65.00	52.50	45.00
IV						
WW	85.00	67.50	85.00	65.00	60.00	50.00
PP	80.00	70.50	62.50	60.00	52.50	50.00

The percentage data for correct recognition responses for sequential positions showed only one consistent pattern. The initial response at the zero-delay condition across all groups in both experiments was relatively errorless and was followed by a drop to a fairly constant recognition level for the remaining positions (see Table 4). Due to the limitations of the percentage data, the sequential positions at zero delay were analyzed with nonparametric tests. The data for WW and PP groups, therefore, were collapsed over experiments and subjected to a series of chi-square tests. The tests revealed no differences between word and picture recognition at the first position. However, both presentation modes showed a significant loss in recognition between Position 1 and 2, with the loss being greater for pictures than for words. That is, for the second position word recognition was superior to picture recognition, a superiority which remained significant throughout the last five sequential positions.

In light of the sequential position analysis, it appears that a new interpretation of the data is possible. The significant loss of recognition following the initial recognition slide seems quite similar to the visual masking effect found in iconic memory (Neisser, 1967). That is, the first slide might be acting as a visual masking stimulus for the iconic memory store, and subsequent recognition performance is poorer due to a visual masking effect. This type of masking effect would not be expected and does not occur at 7.5 and 15 sec. of delay, since the transfer of information from iconic memory to STM should be completed by the time the first cognition slide appears. Furthermore, the results indicate that picture materials may be more subject to this type of interference than verbal materials.

If it can be tentatively assumed that iconic memory is involved in the recognition level at zero delay, the data of Exp. II and III should be evaluated from a new perspective. Thus, the difference in recognition scores for WW and PP at zero delay can no longer be accepted as reflecting purely short-term coding differences because of the damaging effect of the initial recognition slide. Consequently, these data cannot be used as valid reference points in the evaluation of interpolated activity effects. However, in light of the shortcomings of the sequential position data and the resulting tentative acceptance of iconic memory involvement in this experimental setting, it seems necessary to confirm whether iconic memory is indeed a determining factor for zero-delay recognition scores.

Experiment IV, therefore, is designed to further investigate the possible role of iconic memory at zero delay.

EXPERIMENT IV

Characteristic features of iconic memory and the detrimental effects of visual masking on subsequent recall or recognition have been reported in detail by Neisser (1967). Though the present experimental task was not designed to investigate iconic memory, the sequential position data at zero delay display a response pattern implicating the existence of iconic memory in this condition. The problem to be resolved at this point is how to obtain zero-delay recognition data, which are uncontaminated by the visual masking effect. One could reasonably argue that by replacing the visual masking stimulus, i.e., the first recognition slide, with an auditory presentation of the first recognition item, the visual masking effect should be removed from the zero-delay responses. Though it is not clear what additional complications might be introduced by this change, a comparison of the response patterns, especially for the last five positions, should at least provide some indication about the masking effect of the first recognition slide.

Method

The general method of Exp. IV essentially replicated Exp. III, except for one modification: The first recognition item in each sequence was presented orally.

Results and discussion

Recognition scores were again analyzed with respect to correct responses and d′ scores (see Table 5). The results reflect marked changes in both response measures from Exp. III. While the d′ analysis revealed a significant increase in recognition in both groups over delay conditions, $F\ (2,\ 76) = 5.20$, $p < .01$, the previously established word–picture difference did not reappear, $F\ (1,\ 39) = 2.29$. That is, the facilitative effect due to delay was replicated, but the overall word recognition level dropped down to the picture recogni-

TABLE 5

MEAN VALUES FOR BOTH RESPONSE MEASURES AS A FUNCTION OF DELAY CONDITIONS IN EXPERIMENT IV

Cond.	Mean correct responses			Mean d′ scores		
	0	7.5	15	0	7.5	15
WW	6.25	6.95	6.75	2.51	3.46	3.14
PP	5.60	6.80	7.05	2.19	2.71	2.95

Note.—WW = word-word, PP = picture-picture. Column headings 0, 7.5, and 15 indicate the delay in seconds.

tion level, which in turn remained essentially unchanged from Exp. III. The chi-square test for the sequential positions at zero delay also revealed a different pattern from the previous experiments. There was again no word–picture difference for the first position. However, in contrast to the previous experiments, word and picture recognition remained undifferentiated across the six sequential positions, and furthermore revealed a continuous decrease from Position 1 to 6 for both groups. Thus, by replacing the visual with an auditory recognition item, the previous visual masking effect was replaced by an even stronger interference component.

Therefore it appears that in a sequential recognition task of this nature, the initial recognition item, regardless of presentation modality, contributes an interference component for subsequent recognition items. If the first recognition item is presented visually (Exp. II and III), a visual masking effect is obtained. This masking effect, however, appears only at zero-delay conditions and is stronger for pictorial than for verbal stimulus material. If the first recognition item is presented auditorally a qualitatively different interference effect emerges (Exp. IV). In contrast to visual masking, this interference effect takes on a more complex form. At zero delay, d′ scores for word recognition are reduced to the picture recognition level. In addition, recognition shows a consistent decrease across all sequential positions at zero delay, in contrast with the single drop in recognition following the initial item in Exp. II and III. At 7.5 and 15 sec. of delay the interference effect appears only for word groups, while picture recognition remains unchanged from Exp. II and III.

Several parallels can be drawn between the interference effect resulting from the auditory presentation of the first recognition item and investigation of the auditory masking effect (Crowder, 1970; Tell, 1971). Still, an evaluation of the presently obtained auditory interference appears unsatisfactory. The auditorally presented recognition item, for instance, can be considered as an additional verbal stimulus entering the verbal STM, thereby interfering with the storage of the original verbal stimulus material. Hence, word recognition decreases in all three delay conditions by a constant amount (Exp. III vs. IV). A similar decrease does not occur in picture recognition, at least at 7.5 and 15 sec. delay, suggesting that the pictorial material is indeed largely coded nonverbally over short-term retention intervals, and auditory interference does not damage the storage of pictorial material. Zero-delay picture data, however, do not clearly fall within this interpretation. In other words, this particular auditory interference also seems to be damaging to the storage of pictorial material, but only at the zero-delay condition.

GENERAL DISCUSSION

The major finding of this series of experiments lies in the differential recognition patterns as a function of rehearsal and interference conditions. This is especially noteworthy since in a free-recall task this pattern did not emerge. In general, there are three specific trends which will be discussed in detail.

First, the recall and recognition results for verbal material replicate those of earlier studies (Hellyer, 1962; Murdock, 1963) and can be interpreted in terms of the verbal short-term system (Atkinson & Shiffrin, 1968). Because recall performance did not differ for words and pictures in all three groups in Exp. I, picture recall data are also interpreted as reflecting verbal coding. Since response demands in free recall require verbal labeling of words and pictures, the absence of differential coding effects is largely attributed to the specific task demands inherent in a free-recall situation.

The second trend concerns the recognition performance with words and pictures at zero delay. The d′ scores reveal a consistent recognition superiority for words in Exp. II and III. With the introduction of an initial auditory recognition item, d′ scores for the word recognition are reduced to the picture recognition level, which remained unchanged from the previous experiments. However, the sequential position data clearly point out that d′ scores at zero delay cannot be interpreted simply in terms of short-term coding processes. The visual masking effect as well as the damaging influence of the auditory recognition item introduced an interference component for this response condition, such that zero-delay responses cannot be utilized as reference points to evaluate interpolated activities. Despite this shortcoming, some evidence for verbal and visual coding can still be inferred. For instance, in Exp. II and III, the visual masking effect was stronger for pictorial than for verbal material (Neisser, 1967). In Exp. IV, the auditory masking effect reduced d′ scores for the word group significantly, while leaving the picture group relatively unchanged. A similar finding has been reported by Cohen and Grandstrom (1970).

The third trend concerns word–picture recognition scores after retention intervals filled with either verbal interference or rehearsal activity. Since zero-delay response levels have already been shown to be inappropriate reference points to evaluate interpolated activity effects, the effects of interference and rehearsal conditions have to be evaluated with respect to one another. Only the 15-sec. retention interval is common to all experiments and will, therefore, be discussed in more detail. The significant increases for picture recognition over retention intervals, which were obtained in the d′ analyses, therefore can no longer be accepted as increases as a result of visual short-term coding. Instead, the data must be considered as merely reflecting a higher recognition level after 15 sec. of interference or rehearsal than following the masking effect at zero delay. However, it is important to note that picture recognition at 15-sec. delay is essentially unchanged across interpolated activities. In contrast, word recognition at this point clearly displays the facilitative effects of rehearsal over verbal interference. Thus, evidence for differential coding for words and pictures appears quite pronounced.

These data can be readily incorporated into the theoretical framework proposed by Atkinson and Shiffrin (1968). Response patterns for word and picture recall and word recognition consistently replicate previous research investigating verbal STM (Atkinson & Shiffrin, 1968). Picture recognition data furthermore extend the evidence for a predominantly visual STM, as distinct from a predominantly verbal STM, and also provide some indication of visual STM characteristics. Though the present experimental task seemed unsuited to reveal relative efficiences of the two coding systems in an immediate recognition task, the experiments demonstrated that, in contrast to verbal short-term coding, rehearsal and verbal interference do not differentially affect the visual short-term system. Whether the undifferentiated recognition level after the interpolated activity intervals is attributed to the absence of a visual rehearsal mechanism or is attributed to the ineffectiveness of verbal interference on the visual STM remains unresolved.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*, Vol. 11, New York: Academic Press, 1968.

Cohen, R. L., & Grandstrom, K. Reproduction and recognition in short-term visual memory. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 450–457.

Crowder, R. G. The role of one's own voice in immediate memory. *Cognitive Psychology*, 1970, **1**, 157–178.

Hellyer, S. Frequency of stimulus presentation and short-term decrement in recall. *Journal of Experimental Psychology*, 1962, **64**, 650.

Kintsch, W. *Learning, memory, and conceptual processes.* New York, Wiley, 1970.

Milner, B. Visual recognition and recall after right temporal-lobe excision in man. *Neuropsychologica*, 1968, **6**, 191–209.

Murdock, B. C., Jr. Interpolated recall in short-term memory. *Journal of Experimental Psychology*, 1963, **66**, 525–532.

Neisser, U. *Cognitive psychology.* New York: Academic Press, 1967.

Paivio, A., & Csapo, K. Concrete image and verbal memory code. *Journal of Experimental Psychology*, 1969, **80**, 279–285.

Parks, T., Parkinson, S. R., & Kroll, N. E. Visual and auditory short-term memory: Effects of phonemically similar auditory shadow material during the retention interval. *Journal of Experimental Psychology*, 1971, **87**, 274–280.

Posner, M. I. Characteristics of visual and kinesthetic memory codes. *Journal of Experimental Psychology*, 1967, **75**, 103–107.

Posner, M. I., Boles, S. H., Eichelman, W. H., & Taylor, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology*, 1969, **79** (1, Pt. 2).

Potter, M. C., & Williams, M. I. Is there an echo box in pictorial memory? Paper presented at the meeting of the Eastern Psychological Association, New York, April 1970.

Tell, P. M. Influence of vocalization in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 149–156.

Tversky, B. Pictorial and verbal encoding in a short-term memory task. *Perception & Psychophysics*, 1969, **6**, 225–233.

(Received January 26, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 87–91

HEMISPHERE FUNCTION AND COLOR NAMING[1]

STUART DIMOND[2] AND GRAHAM BEAUMONT

University College, Cardiff, Wales

Identical pairs of color stimuli were presented for naming to right-handed and left- or mixed-handed groups of *S*s, and directed to the left hemisphere, right hemisphere, or both hemispheres. Reaction times and blocks to function were recorded. It is seen that response latency is greater when one member of the pair is directed to each hemisphere, indicating either an intrahemispheric enhancement or interhemispheric delay in the analysis of such stimuli. This effect is stronger in right-handed than in left-handed *S*s. No difference is found between the conditions of unilateral presentation. Experimental conditions do not affect the number of blocks, except in an increase in the bilateral presentation condition for left-handed *S*s. The results are discussed in terms of the intrahemispheric independence of visual analysis, the possibility of load sharing in cerebral organization, and the importance of the speech output system.

Failure of the ability to name colors has received some degree of attention in the neurological literature (Critchley, 1965; Kinsbourne & Warrington, 1964). Geschwind and Fusillo (1966), for example, described the case of a patient with a lesion of the left calcarine cortex who failed to name colors correctly and to identify a color when provided with its name. This patient was, however, able to match a color to an object without error, thus ruling out a specific deficit of color perception. They therefore suggested that destruction of the splenium, carrying callosal fibers of the visual regions, dissociated colors seen in the right hemisphere from speech mechanisms of the left. Oxbury, Oxbury, and Humphrey (1969) discussed color anomia and described two varieties—a specific aphasic variety, and a variety specific to the visual modality, which was secondary to the anatomical disconnection between visual input and language centers. Scotti and Splinnler (1970) asserted that in unilateral hemisphere-damaged patients, significant impairment of color perception only is found in patients with right-sided lesions and visual field defects, but they found that such patients were also impaired on other noncolor perceptual tests.

The inability to produce an appropriate color name is a phenomenon not confined to the brain-damaged individual. Bills (1931) described a "block to function" in color naming which he assumed to be symptomatic of fatigue, and which arose in the normal individual, after continuous performance for a protracted period on a color-naming task. Such blocks took the form of a temporary inability to provide the appropriate color name, manifest in an exceptionally long reaction time. A similar effect was provided by Broadbent (1953) and shown to be a general feature of continuous self-paced performance tasks where there are a number of separate stimuli requiring discrete responses, although a considerable period of performance may be required before the effect becomes evident.

The customary approach to the study of blocks is naturally by the performance of a reaction time task, thus providing the investigator with evidence not only about blocks, but about more conventional response latencies to the stimuli employed. A color-naming task was employed in the present experiment for this purpose. The object was to provide information about response times to color stimuli, flashed to the left or right hemisphere, or to both hemispheres, in relation to interhemi-

---

[1] The authors wish to acknowledge the support of Medical Research Council Grant G969/96C for the work reported here, the assistance of Lynne Pritchard in the conduct of the experiment, and also to thank M. Annett for kind permission to use her Handedness Questionnaire.

[2] Requests for reprints should be sent to Stuart Dimond, Department of Psychology, University College, Cardiff, P.O. Box 78, Cardiff CF1 1XL, Wales.

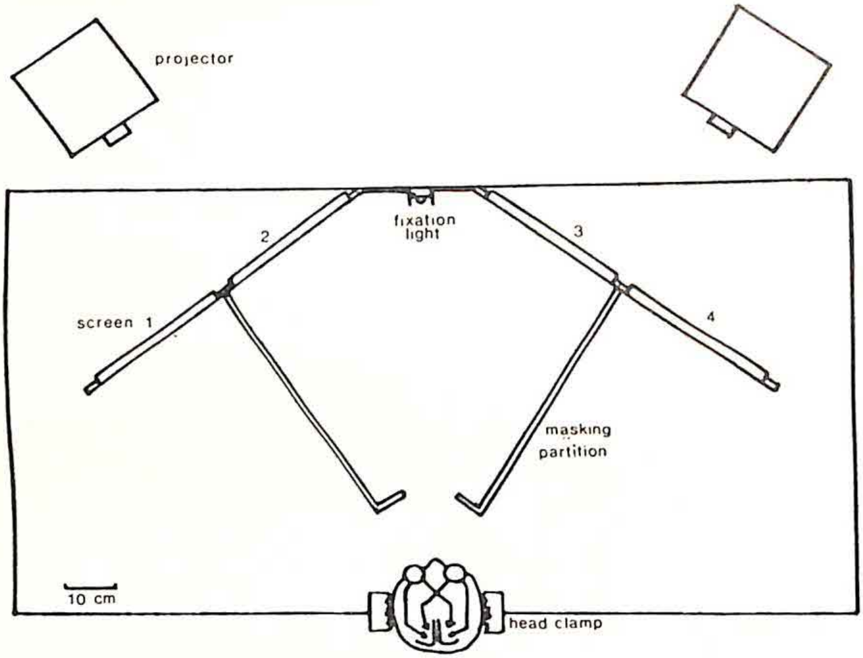


FIG. 1. Diagrammatic scale plan of apparatus.

spheric differences and cooperation in color naming in the normal human brain. It was also a matter of interest to establish how blocks relate to hemisphere function and to question whether the hemispheres might cooperate in order to offset fatigue, by examining whether blocks were produced with equal frequency irrespective of whether the stimuli were flashed to one hemisphere or distributed between the cerebral hemispheres.

METHOD

The *S*s were 36 undergraduates, who may be considered to have been naive, forming two groups of 18 *S*s. One was a right-handed group, achieving preference group scores of 1–8 on the Handedness Questionnaire (Annett, 1970), while the other formed a left- and mixed-handed group, achieving preference group scores of 9–24 on the questionnaire. The experimental conditions for each group were identical.

Stimuli were colored slides produced from a coloring medium (Rank Strand Cinemoid No. 1, 5, 6, 9, 24, 26) in six different colors selected to be of approximately equivalent brightness but adequate discriminability. The stimuli were presented to *S* by means of a divided visual field apparatus, the plan of which appears as Fig. 1. Stimuli were back projected by two automatic projectors (Hanimex Rondette 1200A) on to one of four screens (each 28 × 44 cm., 63 cm. from the *S*) which were arranged together with partitions between the screens and *S*'s eyes so that as *S* gazed at a central fixation light, the material on Screen 1 was directed at the nasal hemiretina of the left eye, on Screen 2 at the temporal hemiretina of the right eye, on Screen 3 at the temporal hemiretina of the left eye, and on Screen 4 at the nasal hemiretina of the right eye. Material on Screens 1 and 2, in the left visual field, was therefore directed to the right hemisphere, and on Screens 3 and 4, in the right visual field, to the left hemisphere. Each trial consisted of a pair of stimuli, of the same color, directed in the left visual field (LVF) condition to Screens 1 and 2, in the right visual field (RVF) condition to Screens 3 and 4, and in the both visual fields (BVF) condition randomly allocated to Screens 1 or 2 and 3 or 4. Thus material in the LVF condition was directed to the right hemisphere, in the RVF condition to the left hemisphere, and in the BVF condition to both hemispheres. There were 120 trials in each condition and the *S* was directed to name the color presented as fast as possible, so activating a voice switch and immediately terminating the trial, which initiated the next. Interstimulus interval was 1.111 sec., and response times were measured from the onset of each stimulus. The order of conditions was randomized, and the fixation of the eyes on the central fixation point was monitored visually by *E* from a peephole adjacent to the fixation light. This appears a reliable method of monitoring, and fixation proved no problem for any of the *S*s.

Blocks were defined as response times greater than the mean plus twice the standard deviation of response time for each *S* in each condition. Mean reaction times were calculated for each *S* and then, together with the number of blocks, combined to form group parameters, and subjected to analysis of variance.

RESULTS

The mean reaction time and number of blocks for each handedness group, analyzed by condition of presentation, are given in Table 1.

Analysis of variance of the mean reaction time in each condition for each *S* showed that there was no significant effect due to the handedness of *S*s, $F\ (1, 612) = 1.70$, $p = .09$, or the color of the stimulus, $F\ (5, 612) = 1.75$, $p = .12$. The effect of the condition of presentation was, however, significant, $F\ (2, 602) = 4.29$, $p = .01$. None of the interactions proved to be significant. Over all *S*s, reaction time was slower for the bilateral presentation condition as compared with the conditions when stimuli were presented to only one hemisphere, BVF–LVF, $t\ (35) = 2.82$, $p < .01$; BVF–RVF, $t\ (35) = 2.34$, $p < .05$. There was no significant difference between the reaction times in the unilateral presentation conditions. If the two handedness groups were analyzed separately, then the left-handed group does not show such an effect, while the right-handed group shows only a significant difference between unilateral presentation to the right hemisphere and bilateral presentation, BVF–LVF, $t\ (17) = 3.04$, $p < .01$. However when the mean results in each condition for the two groups were compared by an unrelated t test, then no significant differences were found between the two groups, showing therefore only a tendency for the effect to be stronger in the right-handed group.

Analysis of variance of the mean number of blocks by condition for each *S* showed no significant effect due to handedness or condition of presentation, and there was no significant interaction. If the handedness groups are compared by an unrelated t test, then no differences are found, except in the number of blocks in the bilateral presentation condition where the left-handed group showed more blocks than the right-handed group, $t\ (34) = 2.35$, $p < .05$.

Blocks to function correlated with reaction time within the two conditions of unilateral presentation for the right-handed group (LVF RT r. LVF blocks = .49, $p < .05$; RVF RT r. RVF blocks = .50, $p < .05$), but this effect was not shown for the left-handed group, or over all *S*s. The handedness measure did not correlate significantly with either the mean reaction time or number of blocks in any condition.

TABLE 1

MEAN REACTION TIME (IN MSEC.) AND NUMBER OF BLOCKS ANALYZED BY HANDEDNESS AND PRESENTATION CONDITION

Group	LVF	RVF	BVF
Right-handed *S*s ($n = 18$)			
RT			
$\bar{X}$	613	635	660
SD	80	107	116
Blocks			
$\bar{X}$	3.39	4.11	3.28
SD	1.72	2.22	1.84
Left/Mixed-handed *S*s ($n = 18$)			
RT			
$\bar{X}$	643	643	661
SD	132	115	101
Blocks			
$\bar{X}$	3.72	3.72	4.28
SD	1.96	2.14	1.41

Errors, although recorded, were too infrequent to permit meaningful analysis, and are consequently not reported here.

DISCUSSION

If *S* is shown two words or two series of digits simultaneously, and these are projected to one hemisphere alone, mutual interference is observed while distribution between the hemispheres leads to increased output (Dimond, 1971; Dimond & Beaumont, 1971). The essential characteristic of such stimuli is that they are different; both require separate analysis and a different response. The present experiment enables us to define something of the limits of this interference as the stimuli are simpler and both members of the pair are identical. If the mutual interference previously observed was a function only of the receipt by a single hemisphere of information from two distinct visual pathways, then it should still be observed here, although the information in the two channels is the same. If, on the other hand, interference arises through the incapacity of the hemisphere to analyze two different simultaneous items of information, then it would not be seen in the present experiment. In fact, it is found that when two stimuli of the same color are projected to a single hemisphere, then the response latency is less than when one stimulus is flashed to one and the other stimulus to the opposite hemisphere.

A channel interpretation of this result would argue that the increase in response time is a function of the increase in the number of active stimulus sources to be monitored in the bilateral presentation condition. However, the information arriving by the channels was the same on every occasion, and unless monitoring of all sources is a requirement of the system (which is improbable as the system could then never benefit from redundancy of information across sources), *S* should have been able to respond as fast as in the unilateral presentation conditions. Some alternative explanation is clearly desirable.

A more acceptable view might be provided in the hypothesis of summation of information along the visual pathways. In the unilateral conditions, one stimulus projects through the ipsilateral pathway, the other through the contralateral pathway, both to the visual cortex of the same hemisphere. At the stage of analysis when the two stimuli must be coordinated, then differing stimuli will result in differing "images" and may possibly set up mutual interference. Identical images, particularly those resulting from similar nondirectional stimuli, may lead to a positive enhancement of the resultant coordinated image. In the case of color, it is as if one color wash were superimposed upon a wash of the same color forming a more intense hue. Such a view suggests that there is an independence of the perceptual trace between one hemisphere and the other and that exposure of the two separate visual mechanisms of the cerebral hemispheres to color stimuli does not lead to potentiation of any subsequent response.

A third interpretation of the increased response latency associated with distributed input is that delay in coordination of the perceptual processes from both hemispheres is an important contributory factor. Although the information conveyed by each hemisphere is the same, the brain is unable to establish this until after the comparison has been made, and it is to this that the momentary delay in response may be due. This view need not be entirely inconsistent with previous findings of an enhancement of response with distributed input in more complex tasks (Dimond, 1971; Dimond & Beaumont, 1971), as the effect of mutual interference with a single hemisphere may be stronger than such a delay in coordination.

Whatever the nature of the explanation, the data relating to the reaction time of responses suggest that there is some degree of initial independence of visual analysis in each half of the brain.

While there is no clear difference between the conditions in which stimuli are presented to either the left or right hemisphere alone, the tendency for reaction times in presentation to the right hemisphere to be faster (as evidenced by the more significant difference between BVF and LVF conditions than between BVF and RVF conditions) is in accord with the clinical literature. The failure to find a true difference may be the result of the importance in the present task of verbal output, allowing the left hemisphere to make some compensation for any real difference in the speed of analysis.

The presence of blocks is commonly assumed to represent a failure of the preceptual mechanisms (Broadbent, 1953). In relation to hemisphere function, the presence of blocks enables us to differentiate between the theory of independent processing in which there is no flexibility in the arrangements between the hemisphere system, and the alternative "suitcase" theory which suggests that the brain offsets fatigue by distribution of visual analysis between the hemispheres, as a suitcase can be interchanged from one hand to the other.

It has been shown, as we have reported, that the two parallel mechanisms of the hemispheres are able to achieve an increased output under certain conditions, when the information load is divided between them. For example, when two parts of a figure, requiring comparison, are flashed to separate hemispheres, the brain works more efficiently than when both parts are flashed to the same hemisphere (Dimond & Beaumont, 1972). It is not clear how the products of analysis within each hemisphere are combined, and while such studies have suggested joint action of the cerebral hemispheres in visual perception, they have not provided evidence of the capacity to switch the load from one hemisphere to the other. The suitcase theory assumes that there is a flexible allocation of load between one hemisphere and the other, so that if the load were to be distributed across the brain, then less fatigue should result than if the load is confined to one hemisphere. In the attempt to test this hypothesis by distributing color stimuli between the hemispheres and observing response blocks it was expected that fatigue would be offset, resulting in a decrease in the number of blocks. That the results show no difference in the number of blocks between the unilateral and bilateral presentation conditions, therefore, provides no evidence for such a theory. It would appear that the hemisphere visual mechanisms are far less flexible than had formerly been supposed, and that there is no mechanism by which fatigue in one hemisphere might be offset by transfer to the other.

It must be considered whether the phenomenon of blocking is some function of a failure of the perceptual mechanisms or of an inability to match colors to their spoken names. Geschwind and Fusillo (1966) and Oxbury, Oxbury, and Humphrey (1969) have suggested that the dissociation of the mechanisms for visual perception from the mechanisms for language may be of prior importance in clinical color anomia, indicating that it is an aphasic rather than a visuoperceptive disorder. In this sense, the blocks reported here are quite unlike those reported previously arising through overloading of the hemispheric visual mechanisms (Dimond, 1970). This view would also be in accord with the subjective difficulty experienced in thought blocking during which it is especially the appropriate name which fails to become available in speech. The failure to find any interhemispheric differences in the number of blocks, indicating that they occur late in the sequence of analysis, at the stage of output, gives some support to this view. However, the finding, in right-handed *S*s on the unilateral presentation conditions, that the number of blocks is significantly associated with mean reaction time within each hemisphere, but not across the hemispheres does suggest some common factor arising from the hemisphere to which the material is initially directed and would not support the hypothesis. It is as if, to some degree, each hemisphere makes a distinct contribution to the features of performance arising from that hemisphere, and that blocks to function are one feature of such a contribution.

There is remarkably little difference between the performance of the right-handed and the left-handed or mixed-handed groups, but one notable effect is a rather smaller degree of facilitation for the left-handed group when the color is presented to one hemisphere alone. Such a finding is in line with proposals (Hecaen & Sauguet, 1971) that the brain is more homogeneous and that hemispheric mechanisms are less clearly differentiated in the left-handed individual. A second effect is the finding that when input is distributed across the hemisphere, then significantly more blocks are observed for the left-handed group. Insofar as the block indicates "fatigue," it would be tempting to conclude that the brain of the left-hander has a reduced overall capacity to sustain performance. However, the true nature of this rather curious observation must await further investigation in view of the wide range of interpretations which might be offered for such a result.

REFERENCES

Annett, M. A classification of hand preference by association analysis. *British Journal of Psychology*, 1970, **61**, 303–322.

Bills, A. G. Blocking: A new principle in mental fatigue. *American Journal of Psychology*, 1931, **43**, 230–245.

Broadbent, D. E. Noise, paced performance and vigilance tasks. *British Journal of Psychology*, 1953, **44**, 295–303.

Critchley, M. Acquired anomalies of colour preception of central origin. *Brain*, 1965, **88**, 711–724.

Dimond, S. J. Reaction times and response competition between the right and left hands. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 513–520.

Dimond, S. J. Hemisphere function and word registration. *Journal of Experimental Psychology*, 1971, **87**, 183–186.

Dimond, S. J., & Beaumont, J. G. Use of two cerebral hemispheres to increase brain capacity. *Nature*, 1971, **232**, 270–271.

Dimond, S. J., & Beaumont, J. G. Processing in perceptual integration between and within the cerebral hemispheres. *British Journal of Psychology*, in press.

Geschwind, N., & Fusillo, M. Colour-naming defects in association with alexia. *Archives of Neurology*, 1966, **15**, 137–146.

Hecaen, H., & Sauguet, J. Cerebral dominance in left-handed subjects. *Cortex*, 1971, **VII**, 19–48.

Kinsbourne, M., & Warrington, E. Observations on colour agnosia. *Journal of Neurology, Neurosurgery & Psychiatry*, 1964, **27**, 296–299.

Oxbury, J. M., Oxbury, S. M., & Humphrey, N. K. Varieties of colour anomia. *Brain*, 1969, **92**, 847–860.

Scotti, G., & Splinnler, H. Colour imperception in unilateral hemisphere-damaged patients. *Journal of Neurology, Neurosurgery and Psychiatry*, 1970, **33**, 22–28.

(Received January 26, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 92–96

TRANSFER OF INFORMATION FROM MANUAL TO OCULOMOTOR CONTROL SYSTEM[1]

RONALD W. ANGEL[2] AND HARRY GARLAND

Stanford University School of Medicine

Eye position was recorded while *S* tracked a visual target. On each experimental run, *S* moved the target actively by means of a joy stick. Target motion was either synchronous with that of the hand or delayed by a fixed amount (.18, .36, .72, or 1.44 sec.) On each control run, target motion was a tape-recorded copy of the pattern generated by *S* during the previous test run. Regardless of the delay between joy stick and target motion, the frequency of saccades was generally smaller during the test run than during the matched control run. The results suggest that information about hand movement is available to the oculomotor system for a significant time following manual performance.

The neural mechanism controlling the hand is quite separate from the oculomotor system, and the latter contains four distinct subsystems which can be excited independently of each other (Robinson, 1968). It follows that several motor systems must be coordinated whenever the hand is moved under visual guidance. What mechanism is responsible for this coordination? Recent experiments (Festinger & Canon, 1965; Steinbach, 1969; Steinbach & Held, 1968) have suggested that there is a transfer of information from one motor system to another. According to this proposal, efferent signals destined for the hand are sent also to the oculomotor system. Eye movements can thus anticipate hand movements, without the delays that occur in sensory feedback.

Among the evidence for this concept is the finding that *S* can track a visual target more accurately when he moves it actively than when it moves independently of his control (Steinbach, 1969). If the improved accuracy is due to a transfer of information from the manual to the oculomotor control system, it is pertinent to ask how long that information persists in useful form. The work described here was designed to study this question.

METHOD

Subjects.—Five male and two female *Ss*, aged 24–54 yr., were paid to participate in the experiment.

Apparatus.—The position of *S*'s right or left eye was recorded by a photoelectric technique (eye-movement monitor, Model SGH/V2, Biometrics, Inc.; headrest, Model 115, Biometrics, Inc.). The visual target was a spot of light displayed on an oscilloscope about 30 cm. from *S*'s eyes. During experimental runs, the position of the target was controlled by *S*, who manipulated a small joy stick (6 cm. long) projecting upward from the table between him and the oscilloscope. As the joy stick was moved to the right or left, the target moved horizontally across the cathode-ray tube in the same direction. The maximal excursion of the target was 9.4 cm., subtending a visual angle of 18°. Throughout each test run, the position of the joy stick was recorded continuously by means of an Ampex (SP-300) tape recorder.

On some test runs, the target responded to movement of the joy stick without any delay. On others, there was a fixed delay between movement of the joy stick and response of the target. A delay of .18 sec. was produced by running the tape recorder at 15 ips and displaying the output of the playback head on the oscilloscope. Delays of .36, .72, and 1.44 sec. were produced by running the tape at $7\frac{1}{2}$, $3\frac{3}{4}$, and $1\frac{7}{8}$ ips, respectively. The tape recorder introduced a fast, low-amplitude ripple causing the target to appear slightly fuzzy, but this appearance was identical during the test runs with delay and the control runs.

Procedure.—For each test, *S* was seated facing the oscilloscope. He was instructed to move the joy stick back and forth irregularly, always keeping the eyes on the target. The pattern of target motion was supposed to be slightly difficult but

---

[1] This research was conducted at the Veterans Administration Hospital, Palo Alto, California. The second author was a doctoral candidate in the Biophysics Program at Stanford University. The statistical analysis was done by Bruce W. Turnbull, Department of Biostatistics, Stanford University School of Medicine.

[2] Requests for reprints should be sent to Ronald W. Angel, Department of Neurology, Veterans Administration Hospital, Palo Alto, California 94304.

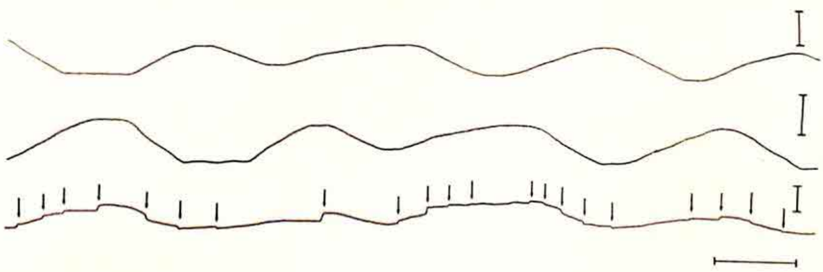


FIG. 1. Tracking of self-moved visual target, with 1.4-sec. delay. (Top line: position of joy stick. Middle line: position of target. Bottom line: position of eye. Movement of target lags 1.44 sec. behind movement of joy stick. Oculomotor response is a combination of smooth pursuit and saccadic jumps [arrows]. During the 10-sec. interval, 21 saccades occur. Calibrations: 5 cm., 20°, 20°, 1 sec.)

not too difficult for the eyes to follow. The standard test run lasted 1 min. and was paired with a control run, during which the joy stick was not used but the pattern of target displacement was an exact copy of the pattern displayed on the test run. This was achieved by playing back the tape-recorded sequence of movements generated by *S*. A complete test consisted of 12 runs: 6 test runs, each paired with a control run.

Immediately after each test run, the tape was backed up to the starting point, a process which required several seconds, and then replayed at the same speed used in recording. During this playback, *S* again was required to keep the eyes directed at the target.

On the first and sixth test runs, there was no delay between movement of hand and target. On each of the other four, a fixed delay was introduced by displaying the output of the tape recorder, as described above.

In order to cancel any effects of practice or fatigue, the four different tape speeds were used in a different order for each *S*. Thus, 7 out of the 24 possible permutations were used. In two cases (S_1 and S_7), the initial run with no delay was omitted inadvertently. In one case (S_7), the run with 1.44-sec. delay was missed.

During each trial, three variables were written out on paper by means of a strip-chart recorder: position of the joy stick, position of the visual target, and position of the eyes. Records were later examined to determine the number of saccadic eye movements per minute of tracking performance. In processing each pair of records, the two target patterns were aligned carefully so that identical 60-sec. periods were taken for study. Thus, any difference in tracking performance could not be attributed to a difference in the target motion during the test and control runs.

One possible source of error in counting saccades was the occurrence of eye blinks. The photoelectric monitor was ineffective during eye closure, but each blink was clearly evident on the record of eye position. Hence the blinks could be counted and their frequency compared under test and control conditions. For each *S*, the total number of blinks recorded during all test runs was compared with the total recorded during all control runs. The ratios (test: control) were: 1:4, 23:36, 1:6, 0:0, 22:10, 26:41, and 5:8. These figures show that five *S*s blinked more often during control runs, one blinked more often during test runs, and one did not blink at all. If some of the saccadic eye movements were not counted because of blinks, this probably occurred more often during control runs, when the majority of blinks occurred. In any case, the number of blinks does not seem adequate to vitiate the results presented below.

Another possible source of error was the problem of classifying some very small deflections on the record of eye position. Large saccadic jumps were easy to recognize, but the smallest were sometimes difficult to distinguish from the background of smooth pursuit motion. In order to minimize the error from this source, two separate tabulations were made: definite saccades and "doubtful" saccades. Whenever a deflection was not large enough or sharp enough to be classified as a definite saccade, it was placed in the doubtful column. In order to determine whether the inclusion of these events might tend to change our conclusions, the results were tabulated a second time, including the doubtful events. Each comparison of test run versus control was examined again, there being 39 such comparisons (6 for each of the seven *S*s minus 3 missed unintentionally). When the doubtful saccades were included along with the definite ones, the algebraic sign of the difference between test and control was unchanged in 37 out of the 39 comparisons. Hence, the results were essentially the same, regardless of whether or not the doubtful events were included in the tallies. For convenience, the numbers presented below include only the definite saccades.

RESULTS

Part of a test run is compared with its control in Fig. 1 and 2. On the test run, *S* was moving the joy stick, and target mo-

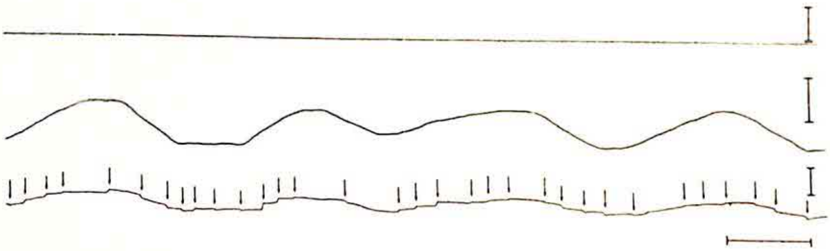


FIG. 2. Tracking identical target motion, reproduced by tape recorder. (Straight line at top indicates that joy stick is stationary. Middle line: exact replica of target motion presented in Fig. 1, reproduced by tape recorder slightly more than 1 min. later. Oculomotor response shows 33 saccades during the 10 sec. Calibrations same as in Fig. 1.)

tion followed after a delay of 1.44 sec. As shown on the bottom line, the oculomotor response was a combination of smooth pursuit and saccadic jumps. During the 10-sec. interval shown, 21 saccades were counted. On the control run (Fig. 2), *S* released the joy stick, and the tape recorder produced an exact copy of the target movements seen during the test run. Thirty-three saccades occurred during the interval shown.

The results obtained from all seven *S*s are presented in the table. An analysis of variance showed that the frequency of saccades during control runs does not change significantly over the delay intervals, F (4, 23) = .46, although there are significant differences between *S*s, F (6, 23) = 31.6, $p < .01$.

The left-hand column shows that when the visual target moved synchronously with the hand, the frequency of saccades ranged from 85 to 158 per min. The corresponding control values ranged from 121 to 332 saccades per min. (The figures presented for S_5 and S_6 are single counts. Each of the others is the mean for two trials, one before and one after the trials with target delay.) In every case, fewer saccades were made on the test run than on the matched control run.

The other columns show the frequency of saccades during runs with various amounts of delay between movement of the joy stick and movement of the visual target. Whether the delay was .18, .36, or 1.44 sec., each *S* showed fewer saccades on the test run than on the paired control run. An analysis of variance was done on the distribution of control-test differences over the delay intervals. This showed that the delay interval had a significant effect, F (4, 23) = 6.11, $p < .01$.

Because of blinking, numerous doubtful saccades, and the differences between *S*s, one should not attach any great significance to the absolute number of saccades counted during 1 min. of tracking. Never-

TABLE 1

NUMBER OF SACCADES PER MINUTE OF VISUAL TRACKING

S	Delay between movement of hand and visual display (sec.)									
	0	C	.18	C	.36	C	.72	C	1.44	C
1	85	121	46	88	64	128	60	94	94	124
2	129	190	128	205	150	193	206	215	165	189
3	152	236	223	283	203	231	297	274	283	298
4	135	252	139	264	201	250	210	205	164	177
5	158	215	159	194	160	207	134	199	157	187
6	114	146	59	152	58	121	157	145	135	167
7	140	332	185	293	124	281	228	255	—	—

Note.—C = control.

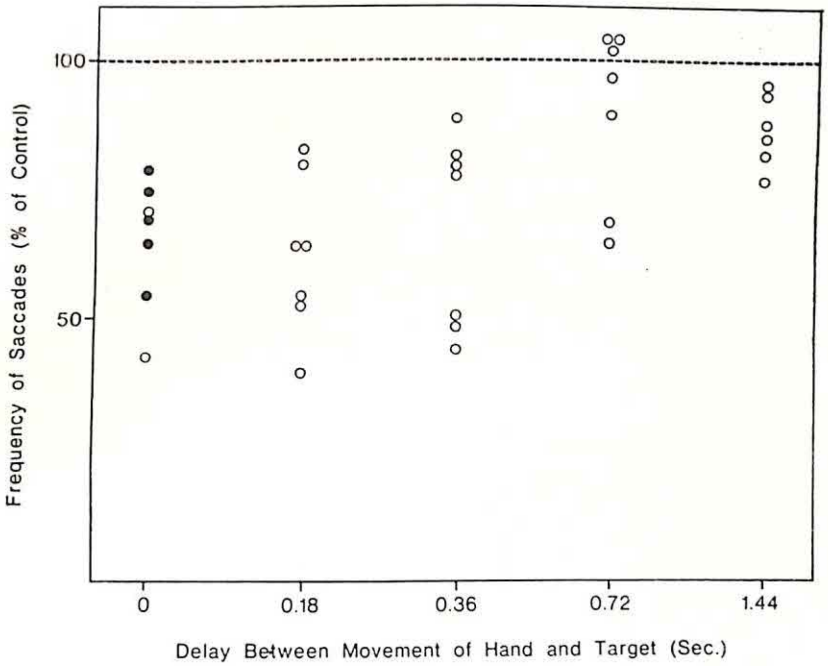


FIG. 3. Frequency of saccades as a function of delay between movement of hand and of visual target. (Each open circle is the number of saccades in 1 min. of visual tracking, expressed as percentage of control. Each filled circle is the mean of two trials, one before and one after the four trials with delayed target motion. Scale of abscissae is logarithmic. The intercept of the best-fitting straight line is 63%, and the slope is 20%/second of delay.)

theless, it may be instructive to consider the relative number, as compared with the matched control run. In Fig. 3, the ordinate of each point is the frequency of saccades, expressed as a percentage of the control. All except three points are seen to fall below the 100% line. In addition, there is an apparent tendency for the relative number of saccades to increase with increasing delay between the hand and target. An analysis of variance, done on the distribution of test/control ratios, showed that these values do change significantly over the delay intervals, F (4, 23) = 6.28, $p < .01$. Since there is no assurance that the test/control ratios are distributed normally, a further analysis was done on the logarithms of the test/control ratios, and these values also were found to change significantly over the delay intervals, F (4, 23) = 5.15, $p < .01$.

DISCUSSION

In previous studies, the frequency of saccadic eye movement has been taken as an index of error in tracking performance (Steinbach, 1969; Steinbach & Held, 1968). Probably a better index would take into account the angular distance between the fovea centralis and the retinal image of the target. For obvious technical reasons, this quantity is much more difficult to obtain than a count of saccadic jumps. However, it is reasonable to assume that saccades occur in response to a misalignment of the fovea and the target image. On that assumption, one may employ the frequency of "corrective saccades" as a crude index of tracking error. A decrease in the number of saccades per minute would thus indicate an improvement of tracking performance.

The results obtained with synchronous movement of hand and target confirm Steinbach's (1969) experiment and also refute a possible objection to that study. In Steinbach's "passive" condition, E moved the target through a pattern of displacements "similar" to the one S produced in the active condition. Since the two patterns were not identical, the observed difference in tracking performance might have been due to subtle differences between them. In the present work, the target followed exactly the same pattern of displacements during test and control runs. Furthermore, each trial lasted 60 sec., as compared with the 10–15 sec. used in Steinbach's experiment. Despite these differences in technique, our results were essentially the same as those reported earlier: All *S*s performed better when the target was under manual control. This finding has been taken as evidence that the oculomotor system has access to outflow information from the system controlling hand movements.

The "outflow" hypothesis is supported by numerous lines of evidence. In monkeys, denervation studies have shown that the absence of vision and proprioception does not eliminate voluntary movements or impair prism adaptation (Knapp, Taub, & Berman, 1963). Normal humans can arrest errors of movement in less than the minimal time needed to process visual or proprioceptive information (Higgins & Angel, 1970). They can also recognize their own errors without external stimuli and in the absence of the signal to which they responded incorrectly (Rabbitt, 1966). Together with the data from eye tracking studies, these findings tend to confirm the reality of a monitored or secondary outflow (von Holst, 1954; Sperry, 1950).

In spite of the evidence for a monitored outflow, this hypothesis is not required to explain the results obtained with asynchronous movement of hand and target. When movement of the visual target lagged behind that of the joy stick by intervals up to 1.44 sec. in length, the eye was still able to follow the target more accurately on a given test run then on the paired control run. Hence it appears that information about movement of the hand is available to the oculomotor system for a considerable time after the movement is executed. During that interval, *S* receives input from the eyes, as well as from the muscles, tendons, joints, and skin, informing him about the movement of the hand. With so much information available, there is no need to postulate that the oculomotor system uses efferent signals from the manual control system. Of course, the data do not exclude the possibility that outflow signals are empolyed in addition to the sensory information, but they certainly do not require that assumption. It is more parsimonious to assume that information about movement of the hand is stored in memory and remains accessible to the oculomotor system for an appreciable length of time.

In a study of adaptation to displaced visual feedback from the hand, Held, Efstathiou, and Greene (1966) found that the correlating mechanism, believed to account for adaptation, cannot handle a feedback signal delayed by as little as .3 sec. This result suggests that the mechanism responsible for prism adaptation is different from that regulating normal hand–eye coordination.

REFERENCES

Festinger, L., & Canon, L. C. Information about spatial locations based on knowledge about efference. *Psychological Review*, 1965, **72**, 373–384.

Held, R., Efstathiou, A., & Greene, M. Adaptation to displaced and delayed visual feedback from the hand. *Journal of Experimental Psychology*, 1966, **72**, 887–891.

Higgins, J. R., & Angel, R. W. Correction of tracking errors without sensory feedback. *Journal of Experimental Psychology*, 1970, **84**, 412–416.

Knapp, H. D., Taub, E., & Berman, A. J. Movements in monkeys with deafferented forelimbs. *Experimental Neurology*, 1963, **7**, 305–315.

Rabbitt, P. M. A. Error correction time without external error signals. *Nature*, 1966, **212**, 488.

Robinson, D. A. Eye movement control in primates. *Science*, 1968, **161**, 1219–1224.

Sperry, R. W. Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative Physiology and Psychology*, 1950, **43**, 482–489.

Steinbach, M. J. Eye tracking of self-moved targets: The role of efference. *Journal of Experimental Psychology*, 1969, **82**, 366–376.

Steinbach, M. J., & Held, R. Eye tracking of observer-generated target movements. *Science*, 1968, **161**, 187–188.

von Holst, E. Relations between the central nervous system and the peripheral organs. *British Journal of Animal Behaviour*, 1954, **2**, 89–94.

(Received January 26, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 97–103

SPEED/ACCURACY TRADE OFF AND PRACTICE AS DETERMINANTS OF STAGE DURATIONS IN A MEMORY-SEARCH TASK[1]

BARRY L. LIVELY[2]

Bowdoin College

In a memory-search experiment, presented items included stimuli belonging to the positive set (which contained one, two, or four consonants or digits) and negative items which were either of the same class as the positive set (both sets were consonants or digits) or of a different class (one set consonants and the other set digits). The slope of the reaction time function for conceptually different negative items was less than for negative items of the same category as the positive set. Practice affected slopes and intercepts for positive and same-category negative items but only the intercept for the negative items of a different category. For each kind of item the speed/accuracy trade off affected only intercept values; its contribution was interpreted as an initial effect on both stimulus encoding and response components, with the effect on response processes decreasing with practice.

It was recently shown (Lively & Sanford, 1972) that membership in a well-learned category can affect the rate at which information is processed in a memory-search task. Using the paradigm developed by Sternberg (1966, 1967, 1969), *S* was given a variable number of consonants or digits to hold in memory, followed by presentation of individual consonants and digits. Reaction time (RT) as a function of the size of the memorized or positive set (PS) was plotted for PS items, NS items (negative items of the same conceptual category as the positive items, i.e., a negative consonant presented when consonants constitute the PS), and ND items (negative items of a different conceptual category from the positive items, i.e., a digit presented when the PS contains consonants). Consistent with Sternberg's (1966) hypothesis of serial and exhaustive search, the functions for Cond. PS and NS were linear and of equal slope (39.9 msec/item and 43.8 msec/item, respectively). The function for Cond. ND was also linear but of lower slope (25.0 msec/item) than the other two conditions. Intercepts were the same in all conditions, suggesting that nonsearch operations such as those required for encoding the presented stimulus and selecting and initiating a response were not responsible for the difference in performance under Cond. ND. In the course of examining alternative explanations for these results, the data were analyzed in terms of *S*'s fastest correct RT under each condition. It was observed that the slope of the function relating size of the PS and fastest RT for Cond. ND (22.8 msec/item) was the same as that for mean RT, while the slopes for Cond. PS (19.0 msec/item) and NS (29.0 msec/item) declined to values approximating that for Cond. ND.

The present experiment was designed to test the generality of the findings for mean and fastest RT under conditions known to affect the slope and/or intercept of the RT × Size of PS function. As in the Lively and Sanford (1972) experiment, stimulus probabilities were equal across Cond. PS, NS, and ND; the design was extended by carrying testing over three sessions instead of one and including (within-*S* and within-session) payoff schedules differentially emphasizing speed and accuracy of responding.

Using conditions comparable to Cond. PS and NS of the present experiment, various investigators have found a reduction in slope and intercept with extensive practice (Briggs & Blaha, 1969; Ross, 1970). This suggests effects in terms of search and nonsearch components. The question here was whether the difference in slope for Cond. NS and ND would be reduced with practice. In addition, the practice variable was included to observe its effect on fastest RT, which had already been shown to be sensitive to size of the PS.

---

[1] This work was carried out when the author was on the faculty of Bowdoin College and was supported by the National Institutes of Health (I-ROJ-MH18154-01). The author is grateful to Michael Bennett for his technical assistance.

[2] Requests for reprints should be sent to Barry L. Lively, Syracuse Psychiatric Hospital, Box 27, University Station, Syracuse, N. Y. 13210.

Further generality would be achieved if the difference in slope for Cond. NS and ND were maintained under different emphases on speed and accuracy. Swanson and Briggs (1969) and Briggs and Swanson (1970) found an effect on the intercept but not on the slope of the function relating size of the PS and RT in a memory-search task manipulating the speed/accuracy trade off. Since the response decision was viewed as being essentially automatic, they assumed the locus of the effect of the trade off to be in the stimulus encoding operation. The present experiment permitted further examination of this interpretation. Unlike their experiment, the present study utilized imbalanced response probabilities, with a negative response being correct two-thirds of the time. If manipulating speed and accuracy has an effect on the postsearch processes of response selection and initiation, an interaction should be evident between speed/accuracy trade off and response type; that is, the difference in RT between Cond. PS and NS should be sensitive to the trade off. This prediction follows from application of Sternberg's (1969) additive factor method. The central assumption of this method is that simultaneous manipulation of independent variables affecting the same stage(s) should produce an interaction, while manipulation of variables affecting different stages should result in additivity of effects. Sternberg has shown that stimulus probability and stimulus type affect different postsearch stages. In the present experiment, those stages are confounded, and the presence of an interaction between speed/accuracy instructions and Cond. PS versus NS would be indicative of an effect on one or more of these postsearch stages. If, however, the speed/accuracy trade off effect is confined to presearch stages, the described interaction should not be in evidence.

The effect of speed/accuracy trade off upon the fastest RT is also of interest. The conditions of the Lively and Sanford (1972) experiment favored accuracy, and the fastest RT was seen to be sensitive to size of the PS, indicating that *S* was in some sense always processing the stimulus. As demands are made for greater speed, it may be the case that *S* would preprogram his response independent of the presented stimulus; given these conditions, the fastest RT should be independent of size of the PS.

METHOD

Subjects.—The *S*s were 12 volunteers (3 of whom were women) who responded to an advertisement. Some of them were Bowdoin College students; others were participants in summer programs for high school teachers. Ages ranged from 17 to 34 yr. The *S*s were paid according to the payoff schedules given below.

Procedure.—The *S* was seated approximately 6 ft. from a white cardboard screen on which slides were projected by a Kodak Carousel 750 projector. At the beginning of a trial, *E* said "Ready" followed 1–3 sec. later by the opening of a shutter which permitted the display of the stimulus. At the onset of stimulus presentation, a photocell embedded at one side of the screen activated a voice key and timer. The trial was terminated by *S* responding "BES" to PS items or "BO" to NS and ND items. These responses were used to equate onset characteristics of positive and negative responses. The shutter was then closed and *E* reported whether *S* was correct and gave his RT to the nearest millisecond.

The consonant and digit stimuli were individually typed in uppercase elite type directly on Kodak Ektagraphic Write-On slides. The stimuli used were the consonants B,F,H,J,K,L,N,R and the digits 1–8. Each of four groups of *S*s had four consonants or digits reserved as PS items, while the NS stimuli were four items of the same category as the PS and the ND stimuli were four items drawn from the other category. Over *S*s, the various consonants and digits were used equally often as PS, NS, and ND stimuli. Size of the PS (one, two, or four items) varied from one trial block to another, with the composition of the PS remaining fixed within each trial block. The three PSs used for each *S* remained the same throughout testing. Composition of the NS and ND stimulus sets remained fixed throughout the experiment.

Within each trial block, the first 8 trials were used for practice, followed by 48 test trials. The test trials were divided equally among PS, NS, and ND conditions; the order of testing conditions was

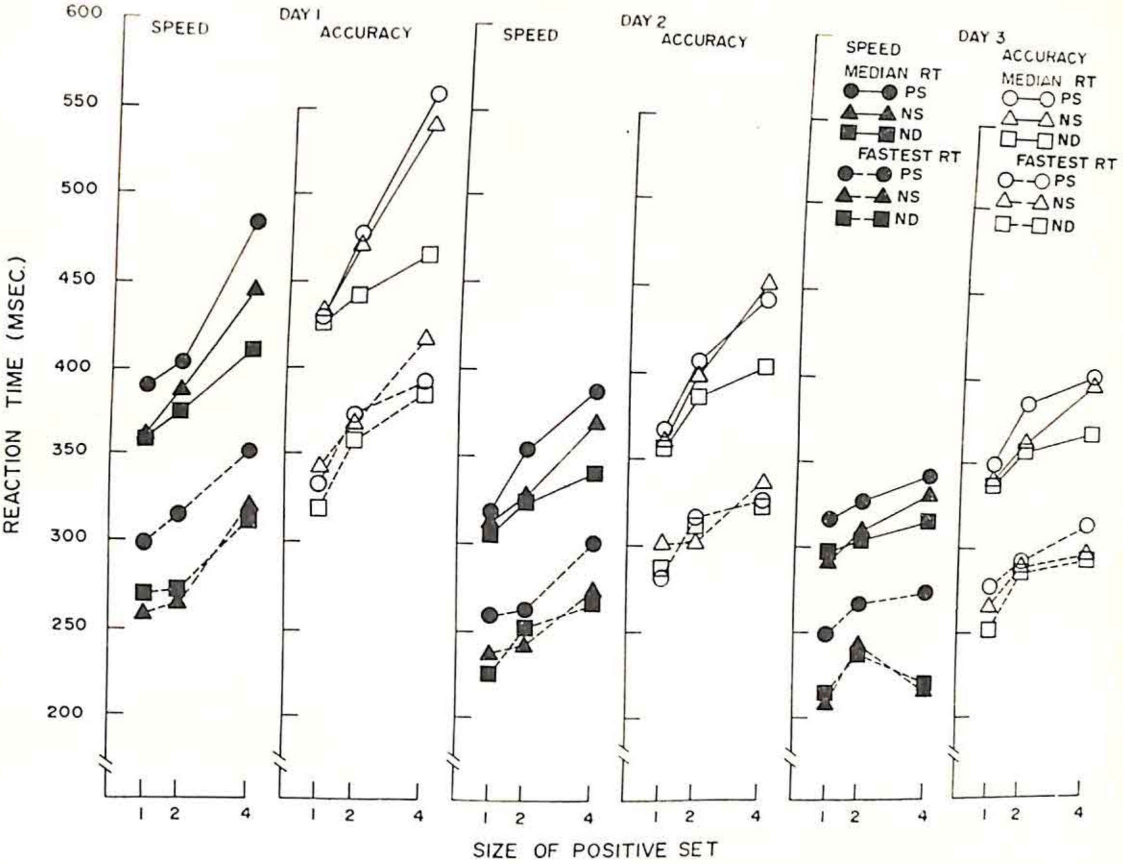


FIG. 1. Median and fastest RT as a function of speed/accuracy trade off, practice, and size of PS.

random with the restriction that no more than four items from the same condition be permitted to occur in sequence.

Each *S* was given six trial blocks on each of 3 successive days. Within each day, half of the *S*s were administered the first three trial blocks under speed instructions, followed by three blocks under accuracy instructions, while the remaining *S*s were tested in the opposite order. The order of speed and accuracy conditions alternated from one day to the next for each *S*. The order in which sizes of the PS were tested was the same under speed and accuracy conditions within each day for a given *S*. The order was rotated from one day to the next for each *S* and was counterbalanced across *S*s each day.

Speed and accuracy instructions.—At the beginning of each set of three trial blocks, *S* was informed whether he was to work for speed or accuracy. He was shown the formula which governed his pay for that set of trial blocks and was given a verbal description of the formula with an example. The formulas used are as follows:

Speed:

$$\text{Pay} = \$.50 + \$.0008[1{,}000 - \overline{RT} - 5(\text{No. Errors})],$$

Accuracy:

$$\text{Pay} = \$.50 + \$.0012[1{,}000 - \overline{RT} - 10(\text{No. Errors})],$$

where $\overline{RT}$ is the mean correct RT over all sizes of the PS and kinds of stimuli (PS, NS, and ND).

The value of speed was emphasized under one condition, while the cost of errors was noted under the other. At the beginning of the second and third sessions, *S* was told the total amount he had earned the previous day.

RESULTS

The median correct RT and fastest correct RT as well as number of errors were determined for each *S* for each condition. Although mean RT was reported in the earlier work (Lively & Sanford, 1972), medians proved to be somewhat more stable here under the varying demands for speed and accuracy; auxiliary analyses have shown that the use of means or medians had no important effect on the interpreta-

TABLE 1

Linear Equations for the Major Reaction Time (RT) Functions in the Experiment

Cond. and day	Speed	Accuracy
	Median	
PS		
1	$352 + 32\,S$	$388 + 43\,S$
2	$302 + 22\,S$	$349 + 25\,S$
3	$310 + 8\,S$	$342 + 16\,S$
Av.	$320 + 21\,S$	$360 + 28\,S$
NS		
1	$332 + 28\,S$	$397 + 36\,S$
2	$292 + 19\,S$	$334 + 30\,S$
3	$282 + 13\,S$	$322 + 19\,S$
Av.	$302 + 20\,S$	$351 + 29\,S$
ND		
1	$340 + 18\,S$	$416 + 13\,S$
2	$297 + 11\,S$	$348 + 15\,S$
3	$290 + 6\,S$	$332 + 10\,S$
Av.	$310 + 12\,S$	$365 + 13\,S$
	Faster	
PS		
1	$282 + 17\,S$	$324 + 18\,S$
2	$242 + 14\,S$	$276 + 14\,S$
3	$246 + 8\,S$	$266 + 12\,S$
Av.	$256 + 13\,S$	$288 + 15\,S$
NS		
1	$232 + 21\,S$	$318 + 25\,S$
2	$222 + 13\,S$	$283 + 13\,S$
3	$223 + 0\,S$	$263 + 9\,S$
Av.	$232 + 11\,S$	$288 + 16\,S$
ND		
1	$252 + 14\,S$	$306 + 21\,S$
2	$220 + 12\,S$	$280 + 12\,S$
3	$223 + 0\,S$	$248 + 13\,S$
Av.	$232 + 9\,S$	$278 + 15\,S$

tion of the data in either the present or earlier experiment. Means for median and fastest correct RTs are displayed in Fig. 1, in which successive pairs of panels give the data for speed and accuracy conditions for each day. An equation for the best-fitting straight line was derived for each set of data; the equations are given in Table 1. These equations are included for purely descriptive purposes; the data were not tested for linearity. In each case the data were pooled over consonants and digits since in no case were consonants and digits found to produce reliably different RTs.

Median correct RT was analyzed separately for Cond. PS, NS, and ND by analysis of variance. Figure 1, Table 1, and supporting analyses show size of the PS to be a highly significant variable under all conditions, $p < .0005$, in each case. Inspection of Fig. 1 and Table 1 suggests that the speed/accuracy trade off was also effective. This was found to be the case under each condition; $p < .025$ for PS, $p < .005$ for NS, and $p < .005$ for ND. The effect of the speed/accuracy trade off was mainly confined to variations in intercept; although slopes of the functions appear to be lower under the speed instructions, a generally reliable effect was not found since the Speed/Accuracy × Size of PS interaction was not significant for either Cond. PS, $p > .10$, or Cond. ND, $F < 1$. An interaction was found approaching significance for Cond. NS, $p < .10$. In all conditions, practice brought about significant reductions in RT; in each case, $p < .0005$. The effect is seen in reduction of both intercept and slope for Cond. PS and NS, while for Cond. ND the effect is confined mainly to the intercept. These observations are confirmed by the Practice × Size of PS interaction; for PS, $p < .0005$, for NS, $p < .005$, and for ND, $.10 > p > .05$. The remaining interactions of Speed/Accuracy × Practice and Speed/Accuracy × Practice × Size of PS produced F ratios of less than one in each condition.

In other analyses of median RT a significant interaction was found between nature of negative stimulus (NS vs. ND) and size of PS, $p < .0005$, replicating the basic result of the earlier experiment (Lively & Sanford, 1972). The relation between the slopes of the NS and ND functions varied according to speed and accuracy instructions, $p < .025$, with the variation in slope being confined to Cond. NS. This effect, however, is small; in Table 1 it is seen that the slope for the average function for Cond. NS under speed instructions is only 9 msec. less than under accuracy instructions.

The difference in slope between Cond. NS and ND decreased with practice. Visual inspection of Fig. 1 and Table 1 show the effect as a reduction in both intercept and slope, the effect being larger for Cond. NS. This conclusion is supported by the significant NS/ND × Size of PS × Practice interaction, $p < .025$, and the NS/ND × Practice interaction, $p < .01$.

The serial exhaustive scanning hypothesis predicts equal slopes for Cond. PS and NS, an expectation borne out by Fig. 1 and Table 1 and the nonsignificant PS/NS × Size of PS interaction, $F < 1$. Furthermore, no overall difference was found between Cond. NS and PS, $p > .05$. The Speed/Accuracy × PS/NS interaction was significant, $p < .05$, the principal effect involving a greater reduction in intercept for Cond. NS under speed instructions. This effect is small and decreases with practice as suggested by the Speed/Accuracy × PS/NS × Practice interaction, $p < .05$. Overall, the difference in RT attributable to the speed/accuracy trade off did not change as a function of practice, $F < 1$. Similarly, the relation between PS and NS showed no change with practice, $F < 1$.

Separate analyses were performed for fastest RT under Cond. PS, NS, and ND. As is evident in Fig. 1 and Table 1, size of the PS was a significant variable, $p < .01$, for each condition, replicating the earlier finding for fastest RT. The speed/accuracy trade off was significant, $p < .01$, in all cases, as was practice, $p < .0005$, in all cases. Except for a significant Practice × Size of PS interaction for Cond. NS, $p < .01$, no other interactions were significant.

In an analysis comparing fastest RTs for Cond. NS and ND, it was found that RT was slightly faster on the average (8 msec.) to ND stimuli, $p < .05$. A significant Speed/Accuracy × NS/ND × Practice interaction, $p < .05$, is indicative of a slight tendency for the difference between NS and ND conditions to decrease with practice under accuracy instructions, while showing a corresponding increase under speed instructions. While significant, the magnitude of the effect suggests it to be rather unimportant. As observed earlier (Lively & Sanford, 1972), the slopes of fastest RT functions were approximately the same under Cond. NS and ND; the NS/ND × Size of PS interaction was not significant, $F < 1$. Although zero slopes were obtained under instructions for speed for both Cond. NS and ND on Day 3, neither the Speed/Accuracy × Size of PS interaction nor the Speed/Accuracy × Size of PS × Practice interaction was significant, $p > .10$ in both cases.

In an analysis comparing fastest RTs for Cond. PS and NS, no overall difference was found between Cond. PS and NS, $p > .10$, while the Speed/Accuracy × PS/NS interaction was significant, $p < .01$. This latter result is seen in Fig. 1 as a tendency for the negative responses to be faster than positive responses under conditions of speed but not accuracy. No other significant interactions were obtained.

Error rates for each condition are given

TABLE 2

PROPORTION OF ERROR RESPONSES AS A FUNCTION OF SPEED/ACCURACY TRADE OFF, PRACTICE, AND SIZE OF POSITIVE SET (PS)

Instruction	Day	Presented item								
		PS: Size of PS			NS: Size of PS			ND: Size of PS		
		1	2	4	1	2	4	1	2	4
Speed	1	.167	.234	.250	.005	.021	.130	.005	.021	.031
	2	.172	.234	.245	.031	.042	.094	.104	.031	.047
	3	.208	.276	.318	.036	.052	.057	.042	.042	.047
Accuracy	1	.031	.068	.100	.021	.021	.042	.005	.042	.021
	2	.068	.100	.135	.021	.010	.057	.000	.021	.010
	3	.104	.078	.115	.010	.021	.036	.005	.016	.010

in Table 2. Analyses of variance reveal systematic effects for Cond. PS and NS, but in no case was significant variation in error rate produced for Cond. ND. The speed/accuracy trade off affected errors for Cond. PS, $p < .0005$, and NS, $p < .05$, as did size of the PS, $p < .025$ and $p < .01$ for Cond. PS and NS, respectively. Practice did not systematically affect error rate in any condition.

DISCUSSION

The capacity for using category information in a memory-dependent search task (as indicated by the difference in slope between Cond. NS and ND) has been demonstrated under conditions which affect the durations of various stages of information processing. The effect was demonstrated under the opposing demands of speed and accuracy instructions which affected nonsearch stages and at various levels of practice which affected search and nonsearch stages alike.

Both the speed/accuracy and practice variables were observed to affect the difference in slope between Cond. NS and ND. The slope of the NS function showed a greater tendency to decline under conditions emphasizing speed or conditions of advanced practice.

The reasons for the differential effect of the speed/accuracy trade off are not clear; part of the explanation may be that under speed instructions and with imbalanced response probabilities *S* tends to preprogram negative responses. Such a strategy would depress the difference between NS and ND slopes, but it would be expected that the fastest RT would be a sensitive index of the strategy. It is only on the third day of testing, however, that the slopes for fastest RT functions under speed instructions reflect such a strategy by evidencing zero slopes.

The finding that practice affected the slope of the function more for NS than ND stimuli may reflect an increase in efficiency of encoding the presented stimulus. Stimulus encoding has been shown here to carry category as well as identity information. Either form of information can be used to reject an ND stimulus as a member of the PS, but only identity information is useful for dealing with the NS stimulus. Confounded with identity is the set of physical features which define an item. Since the same PS, NS, and ND stimuli were used throughout, the role of physical aspects of the stimulus as discriminating features would be expected to increase, thus providing a third source of information. The resulting benefit to performance might be expected to be greater for NS than ND stimuli since the former are more similar to the PS items in the sense of sharing a category.

The speed/accuracy trade off exerted little effect on search time. The locus of the effect is thus restricted to pre- and postsearch stages. Finding a significant Speed/Accuracy × PS/NS interaction suggests an effect of the trade off in the response-selection and initiation stage. The reduction of this effect with practice in the presence of a continuing overall effect of the trade off suggests that initially the durations of both stimulus encoding and response stages are affected by the trade off with the effect on response selection and initiation dropping out with practice.

An interpretation in terms of a presearch stage is compatible with deductions from input sampling models (Fitts, 1966; Stone, 1960). According to these models, the likelihood that a given stimulus was presented changes as information is sampled at a constant rate from the presented stimulus. Sampling is terminated either at the end of a specified interval or after sufficient information has accrued with respect to one stimulus alternative. In either case, the process results in the selection or identification of the stimulus for which the most information has accumulated. The predicted covariation of accuracy and time required to respond was observed here as well as in the work of Swanson and Briggs (1969) and Briggs and Swanson (1970).

Another point is suggested concerning sampling of the stimulus input. It was found here (and by Swanson & Briggs, 1969; Briggs & Swanson, 1970) that intercept values were affected by the speed/accuracy trade off while slopes were not. This implies that the quality of the stimulus representation entered into the search operation is independent of the time required to obtain it. (The further inference that all subsequent processing is independent of sampling time must be tempered by the suggestion that, at least initially, response processes are also affected by the trade off.) Since the identification of the category of the presented stimulus is either dependent on, or simultaneous with, the identification of the name of the presented stimulus, the speed/accuracy trade off, if it only had an effect on the amount of information sampled and not on the process of selecting a response.

would not be expected to modify the difference in slope between Cond. NS and ND.

The fastest RT was found to reflect aspects of stimulus processing and strategies for responding. Except for negative responses under speed instructions on Day 3, the slope of the RT function for each kind of stimulus was positive throughout practice, suggesting that the stimulus was being matched with members of the PS. Response strategies were evident in two forms. In the first 2 days of testing, equal slopes were obtained for positive and negative responses under speed and accuracy instructions. Intercept values for negative responses were more affected by the trade off than intercepts for positive responses. This would seem to mean that response decision and initiation time following search was foreshortened for negative responses under the joint influence of speed instructions and imbalanced response probabilities. On Day 3, these factors apparently produced a gross response strategy of preprogramming negative responses, which was seen as zero slopes for negative responses.

The sensitivity of the fastest RT to a range of variables argues for the continued investigation of its characteristics. It is clear, however, that as a measure it must be interpreted with caution. It follows from the input sampling models (Fitts, 1966; Stone, 1960) that interpretation of differences in mean or median RTs must take into account relative error rates. The counterpart to an error rate for fastest RT is not available, although "fast-guess" models (Ollman, 1966; Yellott, 1967) may eventually be able to provide such an estimate.

REFERENCES

Briggs, G. E., & Blaha, J. Memory retrieval and central comparison times in information processing. *Journal of Experimental Psychology*, 1969, **79**, 395–402.

Briggs, G. E., & Swanson, J. M. Encoding, decoding, and central functions in human information processing. *Journal of Experimental Psychology*, 1970, **86**, 296–308.

Fitts, P. M. Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, 1966, **71**, 849–857.

Lively, B. L., & Sanford, B. J. The use of category information in a memory-search task. *Journal of Experimental Psychology*, 1972, **93**, 379–385.

Ollman, R. T. Fast guesses in choice reaction time. *Psychonomic Science*, 1966, **6**, 155–156.

Ross, J. Extended practice with a single-character classification task. *Perception & Psychophysics*, 1970, **8**, 276–278.

Sternberg, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

Sternberg, S. Two operations in character-recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics*, 1967, **2**, 45–53.

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and Performance II. Acta Psychologica*, 1969, **30**, 276–315.

Stone, M. Models for choice reaction time. *Psychometrika*, 1960, **25**, 251–260.

Swanson, J. M., & Briggs, G. E. Information processing as a function of speed versus accuracy. *Journal of Experimental Psychology*, 1969, **81**, 223–229.

Yellott, S. I., Jr. Correction for guessing in choice reaction time. *Psychonomic Science*, 1967, **8**, 321–322.

(Received February 1, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 104–109

VERBAL REPORT AND VISUAL MATCHING LATENCY AS A FUNCTION OF THE PRONOUNCEABILITY OF LETTER ARRAYS[1]

MAURICE HERSHENSON[2]

Brandeis University

Three experiments are reported comparing performance in response to pronounceable (P) and unpronounceable (UP) arrays of letters. In Exp. I, three groups of *S*s, differing in instructions and in experience with the stimuli, were shown the arrays for a brief duration (19 to 32 msec.). The group that had memorized the stimuli did not show differential performance as a function of the pronounceability of the letter arrays; the other two groups reported more letters correctly for P arrays than for UP arrays. In Exp. II and III, reaction time was measured for "SAME"–"DIFFERENT" response to pairs of arrays presented simultaneously for 100 or 1,000 msec. Trained and naive *S*s were also used in these studies. Response category (S–D) interacted with pronounceability and with number of letters in the array in both experiments. Degree of familiarity with the stimuli and stimulus duration had no effect. Results suggest that the degree of pronounceability of letter arrays plays a role in the retention and report of stimulus information rather than in its first representation.

Why is performance in a recognition task better for arrays of letters which follow the rules of English orthography than for arrays which do not?[3] If the response in the recognition experiment is viewed as the consummation of a series of processes, then the answer may involve (*a*) the reception of stimulus information ("perceptual" explanation), (*b*) the encoding, storage, and retrieval of this information ("memorial" explanation), or (*c*) the availability and production of responses ("response bias" explanation). Perceptual explanations have been proposed in terms of differential "sensitivity" and "expectancy": (*a*) In acquiring reading skills, *S* becomes sensitive to "superforms" organized according to the "auditory–vocal temporal patterns" inherent in the relationship between written and spoken English (Gibson, Pick, Osser, & Hammond, 1962), or (*b*) *S*'s expectations with respect to the stimuli in an experiment act to "tune" the perceptual system to use units of different size for analyzing pronounceable (P) and unpronounceable (UP) stimuli (Aderman & Smith, 1971).

Differential sensitivity has general applicability as an explanation of the pronounceability effect because it develops as a function of reading experience. Expectancy is a concept of limited applicability because expectations are formed in response to the immediate experimental situation, e.g., when similar stimuli are presented in blocks rather than intermixed. In addition, response bias may be eliminated as an explanation because nonverbal responses also manifest the effect (Aderman & Smith, 1971; Gibson et al., 1962). Thus, the pronounceability effect, as a general phenomenon, may be a result of either differential receptive processes or differential memorial processes taking place prior to the response.

The experiments reported studied the role of pronounceability in two different tasks—a recognition task with verbal report of letters seen, and a matching task in which pairs of arrays were judged

---

[1] These experiments were supported, in part, by Grant MH-18092 from the United States Public Health Service. Harry Gray, Linda Handler, and Judith Kroll contributed significantly to this research project.

[2] Requests for reprints should be sent to Maurice Hershenson, Department of Psychology, Brandeis University, Waltham, Massachusetts 02154.

[3] The term "pronounceability" and such phrases as "effect of pronounceability," etc., will be used as shorthand descriptions for these differences. No particular theoretical explanation of the effect is implied by the use of these terms.

"SAME" or "DIFFERENT." The P and UP arrays were randomly intermixed in these experiments to eliminate the possibility that "expectancy" could play a role. To aid in assessing the contribution of storage processes, one group of *S*s was required to memorize the population of stimuli before participating in the experiment. This is a procedure which had been found to be effective in reducing the degree to which the activity of memorial (e.g., differential forgetting) and response processes (e.g., guessing) were represented in verbal responses in a similar recognition experiment (Hershenson, 1969).

EXPERIMENT I

Method

Groups.—There were three groups of *S*s: a Trained-No Guess (T-NG) group, a Naive-No Guess (N-NG) group, and a Naive-Forced Guess (N-FG) group. The T-NG group memorized the population of stimuli used in the experiment. Before beginning the experiment proper, these *S*s scored 95% correct or better on a recall test, given only the first letter in the array (or the first two letters when more than one array started with a particular letter). The N-NG and N-FG groups were naive in the sense that they had not seen the stimuli prior to the experiment. The N-NG group was similar to the T-NG group in two ways: (*a*) On each trial, both groups were shown the stimulus immediately preceding the exposure, and *S* spelled out loud the letters in the array (the "prior knowledge" operation of Haber, 1965); and (*b*) they both were instructed to report the letters they saw after each exposure. If uncertain about what they had seen, they were not to guess but to say "Blank" (the No Guess instruction). The N-FG group was not familiar with the arrays, was not shown the stimulus array prior to the exposure, and was instructed to guess after each exposure even when uncertain about what they had seen (the Forced Guess instruction).

Subjects.—The *S*s were 25 graduate and undergraduate students at Brandeis University. The 5 Trained *S*s were paid $2.00/hr to work for an extended period of time in order to memorize the stimuli. The remaining 20 *S*s were paid at the same rate to serve in the short experiment. These *S*s were randomly assigned to one of the two Naive groups.

Stimuli.—The stimuli were the 50 arrays of four to eight letters each, devised by Gibson et al. (1962): half P and half UP in the sense that P arrays have a high spelling-to-sound correlation and UP arrays have a low spelling-to-sound correlation. The P arrays consisted of an initial consonant spelling followed by a vowel spelling and another consonant spelling, each having a single regular pronunciation. GLOX, BESKS, and QUEESK are examples of four-, five-, and six-letter P arrays, respectively. The UP arrays were constructed from the P arrays by interchanging the positions of the initial and final consonant spellings. For example, the P array GLOX would become the UP array XOGL, the P array BESKS would become the UP array SKSEB.

The stimulus arrays were lettered in black on white paper using a LeRoy lettering stencil No. 0300-290CL with pen No. 4. They subtended a vertical visual angle of .3° and a horizontal visual angle of 2.3° at the maximum. The arrays were exposed in one channel of a Scientific Prototype three-channel mirror tachistoscope (Model GB). The durations for which the stimuli were exposed were determined individually for each *S* in a practice session. A duration was selected for each *S* which would yield correct report of from two to three letters on the average. The actual durations ranged from 10 to 32 msec., with means of 20.4, 25.5, and 19.0 msec. for the T-NG, N-NG, and N-FG groups, respectively. The luminance of both the stimulus and adapting fields was approximately 19 ftl. (65 cd/m^2).

Procedure.—Memorizing the stimuli required a large expenditure of time and energy. Since their status as Trained *S*s would not be altered by multiple exposures of the stimuli, these *S*s participated in two replications of the experiment. The Naive *S*s could not be allowed the same degree of exposure to the stimuli and still remain naive. Therefore, the 10 *S*s in each of the Naive groups participated in only a single replication of the experiment. Thus, 1 *S* in the T-NG group supplied the same amount of data as 2 *S*s in each of the other groups. Nevertheless, the basic datum—percentage of letters correctly reported per cell—was based on the same number of responses in each group.

There were five lists of stimuli, each containing the 50 P and UP arrays in a different random order. A practice list was used on the first day to determine the stimulus durations to be used in the subsequent experimental sessions. The T-NG group saw all five lists, each *S* getting a different random order of lists in one session of about 50 min. The Naive groups saw two lists, randomly selected from the five for each *S*, in one session of about 20 min.

A typical experimental trial started with *S* looking at a fixation field containing two faint horizontal lines which defined the area in which the arrays would appear. The *S*s were instructed to fixate between two dots which bisected these lines. This fixation field remained on during the exposure of the stimulus field and served as a partial mask by decreasing the contrast of the stimuli. For the groups receiving the prior knowledge condition, each trial began with the stimulus array exposed. The channel was extinguished after *S* called out the letters in the array. The *E* then called "Ready," and illuminated the stimulus field for the duration which had been previously determined. The *S* responded by calling out letters in

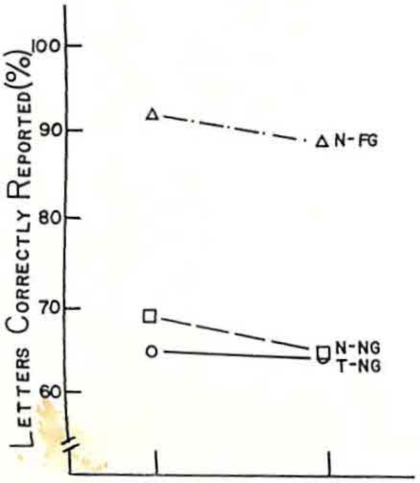


FIG. 1. Percentage of letters correctly reported as a function of pronounceability for each of the three groups in Exp. I.

accordance with the No Guess or Forced Guess instructions.

Results and Discussion

An analysis of variance was computed on 2 arcsin $\sqrt{x}$ transformations of proportions (percentage of letters correctly reported). The N-FG group reported many more letters correctly than either of the other groups, and the N-NG group performed slightly better than the T-NG group, F $(2, 27) = 8.63$, $p < .01$. More letters were correctly reported from P arrays than from UP arrays, F $(1, 27) = 39.23$, $p < .01$. The interaction, F $(2, 27) = 6.38$, $p < .01$, is shown in Fig. 1, where the percentage of letters correctly reported is plotted for P and UP arrays for each group separately. More letters were correctly reported from P stimuli than from UP stimuli by *S*s in both Naive groups, but there was no difference between the two stimulus types for the Trained group (Newman-Keuls test).

Pronounceability entered into the processing when *S*s had either no exposure to the stimuli at all (N-FG group) or had seen the arrays immediately prior to the experimental exposure (N-NG group), and it played a role whether or not *S*s were instructed to guess. Therefore, the superior performance of the N-FG group over that of the N-NG group probably was not a result of differences in the perceptibility of the stimuli, but of the guessing behavior of the N-FG group.

The fact that an effect of pronounceability was not manifest by the T-NG group can be interpreted in two ways: (*a*) The training procedure may have made the UP stimuli into P stimuli, perhaps, by allowing *S* to insert vowels of his own; or (*b*) the training procedure may have eliminated the need for the part of the processing sequence in which pronounceability had its effect. Making P and UP arrays equivalent would eliminate the differential effect without eliminating the process. However, the training procedure has been found to eliminate differential "forgetting" (a memorial process) in a similar experiment (Hershenson, 1969), suggesting that the latter alternative is more probable. If this were the case, the differential effect shown by the N-NG group could have occurred because (*a*) UP stimuli were more difficult to encode than P stimuli, or (*b*) once encoded, UP arrays were not retained in as clear a form as P arrays, or (*c*) UP arrays were harder to decode, i.e., to retrieve from storage, than P arrays.

EXPERIMENT II

If pronounceability enters into the processing in a verbal report task by affecting a component of the information storage process occurring after the first representation has faded, then a task which allows *S* to analyze the first representation directly should yield no differences as a function of pronounceability or of prior experience with the stimulus. One such task might be to require *S* to match two arrays presented simultaneously, to determine whether they are the same (S) or are different (D). The latency of the response in this task could provide an accurate measure of the role of pronounceability and of training.

Method

Stimuli.—The stimulus arrays were drawn from a population of 10 four-letter, 16 five-letter, and 20 six-letter arrays used in the previous experiment. These 46 arrays were used to construct test lists of 60 pairs of arrays consisting of 5 pairs of four, five, and six letters for each of the four conditions defined by a 2 × 2 matrix with S–D on one dimension and P–UP on the other. The 60 pairs were independently selected for each of the four lists and were arranged in random sequence. In addition, a list was constructed containing 18 pairs of arrays of numbers to be used in a practice session. The stimuli were exposed in one channel of the tachistoscope for a duration of either 100 or 1,000 msec. The vertical visual angle of the arrays ranged from .5° to .7°. All other properties of the stimulus arrays were the same as in Exp. I.

Subjects.—The *Ss* were 15 graduate and undergraduate students at Brandeis University. The 5 *Ss* in the Trained (T) group memorized the population of stimulus arrays and were tested as in the previous experiment. The 10 *Ss* in the Naive (N) group had no exposure to the stimuli prior to the experiment. Each *S* in the T group saw all four of the stimulus lists in a different random order, in one session of about 45 min. Each *S* in the N group saw two of the four lists in one session lasting about 20 min. As in the previous experiment, each *S* in the T group produced the same amount of data as 2 *Ss* in the N group. Nevertheless, the basic datum—mean reaction time (RT) per cell—was based on the same number of responses in each group. The *SD* was less than 20% of mean RT for all *Ss* used in this experiment.

Procedure.—The *Ss* were told they would see pairs of pseudowords, one above the other. Their task was to indicate whether the pairs were the same or were different by releasing one of two microswitch keys as quickly as they could. All *Ss* used their preferred hand for the different response. Each trial began with *S* looking into the tachistoscope at a fixation channel which contained two faint lines demarking the areas in which the pairs of arrays would appear. The *S* was instructed to fixate between two faint dots which bisected the two lines. This fixation field was off when the stimulus field was on. Prior to presentation of the stimulus, *E* called "Hands down," then "Ready," and then initiated the exposure of the stimulus. RT was measured from the onset of the stimulus field by a Hunter counter/timer (Model 1520). The two stimulus durations were presented in blocks counterbalanced across *Ss* in the N group and with *Ss* in the T group. Practice was given before each block.

Results

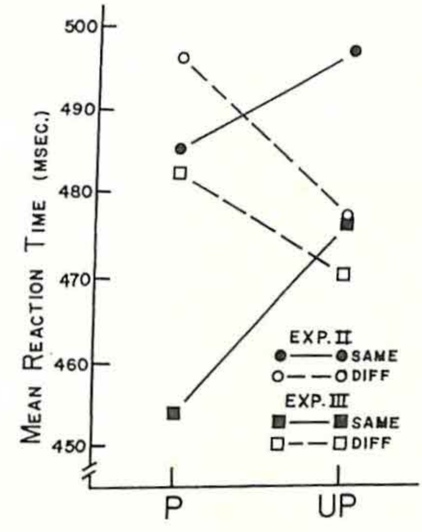


FIG. 2. Mean RT as a function of pronounceability for "same" and for "different" responses in Exp. II and Exp. III.

Analyses of variance were computed on mean and median RT per *S* per cell for correct responses, and on a reciprocal transformation of the mean RTs. The *F* values reported are from the latter analysis although the results were the same for all three. The T group responded more quickly than the N group by about 80 msec., $F\ (1,\ 18) = 8.33,\ p < .01$. This was the only significant main effect. Figure 2 shows mean RT as a function of pronounceability for the two response categories (S–D), $F\ (1,\ 18) = 19.40,\ p < .01$. The UP-D condition was faster than all of the others and significantly different from the P-D and the UP-S conditions (Newman-Keuls test). Figure 3 shows mean RT as a function of number of letters in the array for the two response categories, $F\ (2,\ 36) = 23.31,\ p < .01$. For S arrays, RT was faster the shorter the array; for D arrays, RT was faster the longer the array. RT to six-letter-D arrays was as fast as to four-letter-S arrays; RT to four-letter-D arrays was almost as slow as to six-letter-S arrays (Newman-Keuls test). In general, the number of errors made per condition followed the mean RTs for those interactions—the greater the mean latency, the greater the number of errors.

EXPERIMENT III

Since the matching task did not yield an interaction between pronounceability and

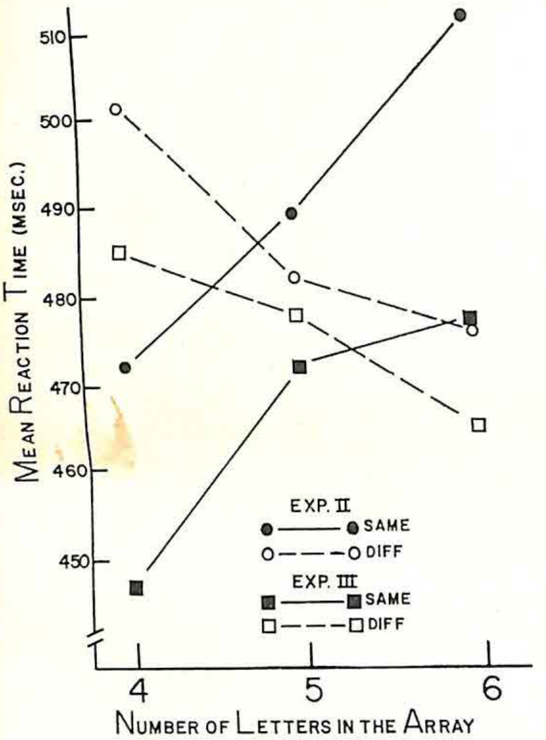


FIG. 3. Mean RT as a function of the number of letters in the array for "same" and for "different" responses in Exp. II and Exp. III.

training, another experiment was performed to replicate that finding. In order to increase the number of responses per cell without simultaneously increasing the Naive group's exposure to the stimuli, the population of stimuli was enlarged from 56 to 180 arrays.

Method

Stimuli.—The population of stimuli was enlarged to 180 four-, five-, and six-letter arrays by constructing new arrays according to the procedure outlined by Gibson et al. (1962). The stimulus duration was 100 msec.

Subjects.—The *S*s were eight undergraduate students at Brandeis University. They were divided into two groups: a Trained (T) group and a Naive (N) group. Trained *S*s had memorized the arrays and were tested as in the previous experiments. As an additional check, the arrays were exposed for a very short duration in the tachistoscope, and *S* had to recall all of the letters in the array after each exposure to a criterion of over 90% correct.

Procedure.—The instructions and general procedures were the same as in Exp. II. The stimuli were divided into two lists of 90 stimulus pairs each with an equal number of S and D pairs in random sequence. No array appeared twice on a list (except in a single S condition). The two lists were presented in an order counterbalanced over *S*s. Extra stimulus pairs were available at the end of each list so that errors could be replaced with additional trials to provide an equal number of responses per cell. Half of the *S*s in each group used their preferred hand for the S response and half used their preferred hand for the D response.

Results and Discussion

Preliminary analyses showed no differences attributable to preferred hand use and no interactions; therefore, the data were collapsed over this variable. Analyses of variance were computed on mean and median RT per *S* per cell for correct responses, and on a reciprocal transformation of the mean RTs. The *F* values to be reported are from the latter analysis.

The T group responded more quickly than the N group by about 100 msec., $F\ (1, 6) = 20.94$, $p < .01$. Response type (S vs. D) interacted with pronounceability, $F\ (1, 6) = 9.21$, $p < .05$, and with number of letters in the array, $F\ (2, 12) = 8.97$, $p < .01$. The P-UP × S–D interaction is shown in Fig. 2. Correct RTs to S stimuli were faster for P arrays than for UP arrays; the other points did not differ (Newman-Keuls test). Figure 3 shows mean RT as a function of the number of letters in the array for the two response types. The latency of a correct response to stimuli which were the same increased as the number of letters in the array increased while the latency for a correct response to stimuli which were different decreased as the number of letters in the array increased. Four-letter-S pairs yielded shorter response latencies than any of the others (Newman-Keuls test). The number of errors, although quite small, followed the pattern of the two significant interactions.

The data from the matching experiments are difficult to interpret with respect to the unit of analysis. If P arrays were processed using a larger unit of analysis (e.g., spelling patterns) than that for UP arrays, the latency of response to P arrays should have been shorter than to UP arrays, whether S or D. This was not the case. RT to P arrays was shorter than to UP arrays when the stimuli were the same but not when they were different—RT to P-D did not differ from RT to UP-S or UP-D. Clearly any statement about units of analyses based on P-UP differences could not be supported by these findings.

A similar paradox exists in the interaction of response type with array size. Note that this variable is *not* "number of letters in the array which are different" but "total number of letters in the array." The increase in RT to S stimuli as the number of letters in the array increased suggests that *S*s were processing letter by letter. However, it is not clear why the opposite should be true for D stimuli. As a consequence of the procedures used to construct the arrays, it is likely that all of the letters in any given pair were different and, moreover, that the first letters were different. Thus the likelihood of hitting a different element should be independent of array size, and there should have been no differences in response latency for D stimuli as a function of array size. Moreover, even if different units were the basis for the differences in the functions, a process would be needed to determine which unit to use, prior to determining whether the arrays were the same or different.

General Discussion

Taken together, the three experiments suggest (*a*) that some process was present in the visual recognition-verbal report task that was not present in the RT task, and (*b*) that this process was eliminated from the sequence in the verbal report task by the training procedure. "Encoding," "storage," and "decoding" are three processes which may play such a role—their function could depend upon the relative degree of pronounceability of the stimuli, and they may play a minimal role in the matching task. If this was the case, then we may conclude that pronounceability plays a role in the retention and report of stimulus information rather than in its first representation.

There were two important variables in the RT experiments which did not yield significant effects: Exposure duration yielded no main effect or interaction (Exp. II) and, while training improved the overall performance in both experiments, there were no interactions. Thus, *S*s were able to perform the matching task equally as well whether the stimuli were novel or familiar, and whether they had a long or a short look at them.

REFERENCES

Aderman, D., & Smith, E. E. Expectancy as a determinant of functional units in perceptual recognition. *Cognitive Psychology*, 1971, **2**, 117–129.

Gibson, E. J., Pick, A., Osser, H., & Hammond, M. The role of grapheme-phoneme correspondence in the perception of words. *American Journal of Psychology*, 1962, **75**, 554–570.

Haber, R. N. The effect of prior knowledge of the stimulus on word recognition processes. *Journal of Experimental Psychology*, 1965, **69**, 282–286.

Hershenson, M. Stimulus structure, cognitive structure, and the perception of letter arrays. *Journal of Experimental Psychology*, 1969, **79**, 327–335.

(Received February 1, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 110-113

IMMEDIATE, GENERALIZED, AND ENDURING EFFECTIVENESS OF PUNISHMENT AND RESPONSE PREVENTION OF HUMAN AVOIDANCE RESPONDING[1]

RAYMOND W. DRAKE, JR.,[2] AND PHILIP A. MEYER[3]

University of South Dakota

Several aspects of effectiveness were examined for punished extinction and response prevention of avoidance responses (ARs) relative to each other and normal extinction. The influence of associating a visual cue with these treatments was also examined. Thus six groups ($n = 12$) of undergraduates were trained to avoid a noxious air blast by pulling a lever during a 3-sec. warning interval signaled by one of two lights. Three phases of extinction were employed to investigate the several aspects of effectiveness for each Treatment × Cuing combination. Results indicated that punishing the AR led to greater immediate suppression than response prevention which, in turn, was more suppressive than normal extinction. Cuing the treatments resulted in generally less immediate resistance to extinction. Subsequent extinction phases indicated that only the uncued punishment procedure yielded any significant generalized and enduring effects. The implications of the present design and results were discussed.

The persistent nature of a learned avoidance response (AR) has been suggested by its well-documented high resistance to extinction (cf. Beecroft, 1967). This aspect of avoidance behavior has led to investigations of other procedures with which to suppress and eliminate learned ARs. The present study is concerned with the relative effectiveness of two such procedures, punished extinction and response prevention.

In comparing the effectiveness of punished extinction and response prevention, two general design factors should be emphasized. First, the notion of effectiveness implies several aspects of behavioral analysis. While studies of procedures designed to facilitate extinction of ARs generally include a normal extinction condition as a base-line control, little attention has been given to effectiveness in terms other than immediate suppressive effects. The significance of enduring effectiveness has already developed with respect to response prevention (Benline & Simmel, 1967; Polin, 1959). In addition, a traditional issue with punishment has been the permanence of its suppressive effects on an instrumental appetitive response (e.g., Boe, 1966; Estes, 1944). Quite another characteristic of effectiveness involves generalizability, i.e., the extent to which the suppressive effects of a procedure influence similar performance in other stimulus environments.

Most previous studies on preventing the AR in extinction have, in one way or another, presented *S* with discriminable cues distinguishing between training and prevention, thereby signaling the impossibility of performing the AR (e.g., placing a barrier between the two compartments of the shuttle box). Studies of punished extinction, on the other hand, have typically used an uncued procedure; i.e., no stimuli have been uniquely paired with punished extinction. Thus, the present investigation also examined the extent to which cuing or not cuing these procedures influenced their effectiveness.

The present study, then, evaluated the immediate, generalized, and enduring effectiveness of the two suppressive procedures of punished extinction and response prevention relative to each other and to normal extinction. Furthermore, these comparisons were achieved under either cued or uncued conditions for each treatment. These conditions constituted a six-group study forming a 3 × 2 factorial design to include treatments and cuing as major variables between *S*s.

---

[1] This article is based on a dissertation by the first author, supervised by the second author, and submitted to the Graduate School of the University of South Dakota in partial fulfillment of the requirements for the doctoral degree.

[2] Now at Wisconsin State University—Platteville.

[3] Requests for reprints should be sent to Philip A. Meyer, Department of Psychology, University of South Dakota, Vermillion, South Dakota 57069.

METHOD

Subjects.—The *S*s were a total of 72 male and female undergraduates at the University of South Dakota. Each *S* received $2.00 or extra credit in Introductory Psychology for participating.

Apparatus.—The apparatus consisted of four major units: (*a*) a dental chair in which *S* was seated; (*b*) a pressurized air delivery mechanism for presenting the noxious stimulus; (*c*) switching and timing circuitry which controlled stimuli presentation, interstimulus and intertrial intervals (ISI and ITI), and latency recording; (*d*) the lever-pull manipulandum and response-prevention device. Excepting the last unit, the apparatus has been described previously (Meyer, 1970). The manipulandum was housed in a 91.37 × 60.91 × 91.37 cm. gray frame presenting *S* with a slanted table top. Atop the frame was mounted a black box containing four different colored, 7½-w. signal lights, and 45.69 cm. from *S*'s upper torso a lever which could traverse an excursion of 5.08 cm. On prevention trials, the lever could move only 3.17 mm. (an attempted response) due to a solenoid operation that was not visible to *S* and produced auditory cues similar enough to those of a dummy solenoid, which operated on all other trials, to further hinder detection.

Procedure.—After being seated in the chair and positioned in front of the frame, *S* was handed instructions which indicated that an annoying puff of pressurized air would occasionally be directed behind the right ear and that during this first brief part of the study physiological recordings would be made from *S*'s forearms strapped atop the frame. Also, *S* was directed to use an earplug. After adjusting the chair and securing *S*'s head to the headrest with an Ace bandage, *E* positioned the air nozzle at a 45° angle pointed at the right ear. The forearm straps served to retain *S*'s forearms in a fixed position, but along with dummy wiring also furthered the illusion of physiological recording which in fact did not occur. Three signaled 4-sec. air blasts were then presented to *S* associated with each of two signals (white and blue lights). White noise was employed throughout. The ISI and ITI were 3 and 4 sec., respectively.

Following these six adaptation trials, *E* released *S*'s right forearm and presented additional written instructions, which indicated that *S* could prevent the air blast. Each trial during training included a 3-sec. white or blue warning signal light (alternated every 5 trials) followed by a 40-psi 4-sec. air blast unless *S* pulled the lever during the ISI. The ITI was 5 sec. and 40 trials were administered. For the first 10 training trials, a response during the air blast terminated it. If *S* made not more than one AR by Trial 10, *E* said: "Try pulling it sooner, (*S*'s name)." After Trial 10, no escape contingency was programmed. Any *S* not making four ARs by Trial 30 was dropped from the study; nine additional *S*s were so dropped and are not included in the design nor in any of the results.

A total of 72 *S*s did successfully complete training and were randomly assigned to one of six groups with the restriction of counterbalancing sex. The groups were differentiated in terms of extinction procedure (normal, punished, or response prevention) and whether an additional cue was present. For normal extinction trials, the air blast was never presented following the signal. For punished extinction, the air blast followed the signal only if an AR occurred during the signal. For response prevention, the air blast was never presented and the lever was immobile as described earlier. Thus six groups were constituted for the three phases of extinction: (*a*) cued normal (CN), (*b*) uncued normal (UN), (*c*) cued punished (CP), (*d*) uncued punished (UP), (*e*) cued response prevention (CRP), and (*f*) uncued response prevention (URP). Extinction was carried out in three phases lasting 20, 20, and 15 trials, respectively, or until *S* did not respond on 5 consecutive trials. Phases 2 and 3 followed directly after the final trial of the previous phase. Response prevention and punishment occurred in Phase 1 only; otherwise normal extinction was administered. The warning signal used in Phase 1 was the training light not presented on Trials 36–40. In the cuing conditions, a green light was presented along with the warning signal during Phase 1 only. In Phase 2, the signal was as on Trials 36–40 to test for generalization of suppression in Phase 1. The Phase 1 signal was again presented in Phase 3 to test for recovery of suppression.

RESULTS

During training, all *S*s reached the criterion of four consecutive ARs within the first 24 trials of the 40-trial training period. An analysis of the mean number of ARs on the last 15 trials (i.e., after all *S*s had reached criterion) indicated no significant differences between cuing conditions ($F < 1$), suppression treatments, $F(2, 66) = 1.89$, $p > .25$, or the interaction ($F < 1$).

Group mean number of trials to reach criterion of five consecutive nonresponses (TC_5) in each extinction phase are presented graphically in Fig. 1 as a function of treatments with cuing conditions plotted separately. The labeling of the

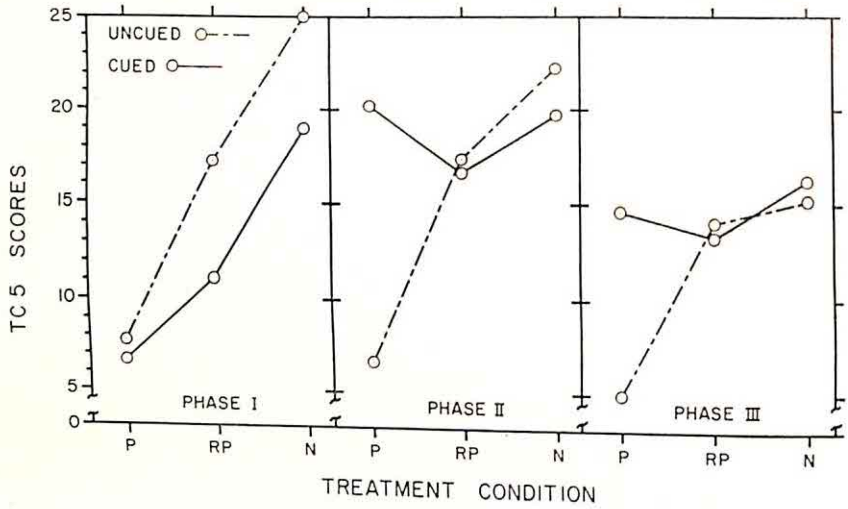


FIG. 1. Group mean trials to criterion scores in the three extinction phases for punishment (P), response prevention (RP), and normal (N) treatment conditions. (The cued and uncued groups are shown separately.)

abscissa does not represent any continuity between treatments.

An analysis performed on the TC_5 scores represented in the Phase 1 portion of Fig. 1 revealed significant effects of cuing, F (1, 66) = 8.98, $p < .005$, and treatment, F (2, 66) = 34.62, $p < .0005$, without a significant interaction, F (2, 66) = 1.39, $p > .25$. Subsequent comparisons of treatment conditions using the Tukey A procedure indicated significantly ($p < .01$) faster extinction for punishment than response prevention which, in turn, extinguished significantly sooner than normal extinction.

Analysis of the Phase 2 data revealed a significant Cuing × Treatment interaction, F (2, 66) = 7.79, $p < .001$, and treatment effect, F (2, 66) = 5.65, $p < .01$. The cuing effect approached significance, F (1, 66) = 3.80, $.05 < p < .10$. Tukey A comparisons indicated significantly ($p < .01$) faster extinction of the UP group relative to any of the other groups, which were not significantly different from one another.

Although the CP and UP groups both extinguished quickly in Phase 1, the CP group took longer to reach criterion in Phase 2 than the UP group. A two-factor analysis including these two groups over Phases 1 and 2 yielded a significant Group × Phase interaction, F (1, 22) = 26.76, $p < .0005$, as well as significant main effects for group and phase, Fs (1, 22) = 27.74 and 19.38, respectively, ps $< .0005$. Tukey A comparisons supported the significant ($p < .01$) increase in TC_5 scores from Phase 1 to 2 for the CP group and, in Phase 2, the slower extinction of the CP group than the UP group. Similar analyses of the response prevention and normal extinction groups failed to yield statistically significant support for any phase effect or Group × Phase interactions.

Analysis of the Phase 3 data yielded significant Cuing × Treatment interaction, F (2, 66) = 4.34, cuing, F (1, 66) = 4.79, both ps $< .05$, and treatment, F (2, 66) = 4.80, $p < .025$, effects. As with Phase 2, Tukey A comparisons corroborated the significantly ($p < .01$) lower TC_5 scores for the UP group below any of the other groups, which did not differ significantly from one another. The apparent depression of TC_5 scores during Phase 3 is simply attributable to the five fewer trials used in this phase as a maximum score.

DISCUSSION

The present study was primarily concerned with the relative effectiveness of three methods for eliminating a learned AR. Considering only the immediate effectiveness of the suppression treatments (Phase 1), punishing the AR in extinction led to greater suppression than either response prevention or normal extinction. Preventing the AR was, however, more effective than the normal extinction procedure. Moreover, associating a visual cue with the treatments resulted in generally less resistance to extinction than when these treatments were administered without a cue.

Based on these findings, the expected conclusion may be reached that both punished extinction and response prevention are more suppressive than normal extinction of an AR. In addition, this investigation demonstrated, for the first time experimentally, differential effectiveness between punishment and prevention. While punishment resulted in rapid response suppression for all of the *S*s, response prevention produced a somewhat more variable performance pattern. Response prevention led to abrupt and early cessation of responding in the majority of *S*s, somewhat "hesitant" cessation in others, and in still others, no cessation at all. A total of five *S*s in Groups CRP and URP attempted to respond on every trial while response prevention was in effect. This variable performance pattern was measured in terms of attempted responses and seems consistent with anecdotal evidence suggesting that *S*s do attempt to perform the prevented AR (e.g., Polin, 1959).

As a possible interpretation of the enhanced suppressiveness of treatments by cuing, the notion of generalization decrement may be utilized. This concept has been considered in general analyses of extinction performance (e.g., Capaldi, 1967). Moreover, a previous study of cued vs. uncued punishment (Moyer, 1955) has obtained similar findings.

Resistance to extinction in Phases 2 and 3 demonstrated the superior generalized and enduring effectiveness of uncued punishment relative to the other conditions. While cuing punishment led to greater immediate suppressiveness, as compared to prevention and normal extinction, this superiority was absent during both the generalizability and durability phases. With regard to response prevention, no strong evidence was obtained for any more generalized or enduring effectiveness than normal extinction. This aspect of the current results is consistent with the findings of other response prevention studies which have questioned the durability of the prevention procedure. Unlike the case with punished extinction, there was no reliable effect of cuing on the subsequent effectiveness of prevention during Phases 2 and 3. This pattern of results with response prevention highlights the reservations which have been expressed regarding its potential as an AR-eliminating procedure (Coulter, Riccio, & Page, 1969; Page, 1955).

REFERENCES

Beecroft, R. S. Emotional conditioning. *Psychonomic Monograph Supplements*, 1967, **2**(4, Whole No. 20).

Benline, T. A., & Simmel, E. C. Effects of blocking of the avoidance response on the elimination of the conditioned fear response. *Psychonomic Science*, 1967, **8**, 357–358.

Boe, E. E. Effect of punishment duration and intensity on the extinction of an instrumental response. *Journal of Experimental Psychology*, 1966, **72**, 125–131.

Capaldi, E. J. A sequential hypothesis of instrumental learning. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 1. New York: Academic Press, 1967.

Coulter, X., Riccio, D. C., & Page, H. A. Effects of blocking in instrumental avoidance response: Facilitated extinction but persistence of "fear." *Journal of Comparative and Physiological Psychology*, 1969, **68**, 377–381.

Estes, W. K. An experimental study of punishment. *Psychological Monographs*, 1944, **57**(3, Whole No. 263).

Meyer, P. A. Role of an unavoidability procedure in eliminating avoidance behavior with humans. *Journal of Experimental Psychology*, 1970, **86**, 337–340.

Moyer, K. E. A study of some of the variables of which fixation is a function. *Journal of Genetic Psychology*, 1955, **86**, 3–31.

Page, H. A. The facilitation of experimental extinction by response prevention as a function of the acquisition of a new response. *Journal of Comparative and Physiological Psychology*, 1955, **48**, 14–16.

Polin, A. T. The effects of flooding and physical suppression as extinction techniques on an anxiety motivated avoidance locomotor response. *Journal of Psychology*, 1959, **47**, 235–245.

(Received February 1, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 114–123

INTERACTION OF VISION AND TOUCH IN CONFLICT AND NONCONFLICT FORM PERCEPTION TASKS[1]

EPP A. MILLER[2]

George Washington University

Three studies explored the importance of the amount of tactual information in visual–tactual interaction. Experiment I used a form conflict task. The *S*s felt an object while viewing its distorted shape. Despite the enhancement of tactual information by (*a*) active touch and (*b*) several trials, vision dominated touch as in previous studies. In Exp. II, two different objects were simultaneously presented, one to vision and one to touch, but without conflict. Here, touch was independent of vision. The striking contrast in the results of these two very similar tasks pointed to the importance of conflict itself. Accordingly, Exp. III analyzed a form conflict. The finding that only the expectation of dealing with a single object induced visual dominance emphasizes a cognitive component in visual–tactual form perception.

Studies of sensory conflict provide an approach to the problem of how the information from our diverse perceptual modalities is integrated into a unitary perception of the world. For example, when *S*s wearing distorting prisms that displace the visual image of a target are asked to point to the target in space, they point to the visual image, i.e., to a different location than where their proprioceptive and kinesthetic sensations would indicate. In such a conflict between vision and touch, vision initially dominates touch, as if vision were used to the exclusion of touch (Hay, Pick, & Ikeda, 1965; Over, 1966). As exposures to the distortion are repeated, *S*s' responses become more veridical, suggesting that the tactual information can nevertheless be taken into account (Held, 1965; Rock & Harris, 1967). One explanation of such adaptation to sensory rearrangement is that the amount of information processed by each modality determines how the discrepancy will be resolved (Freedman, 1968; Howard, Craske, & Templeton, 1965; Moulden, 1971; Wallach & Karsh, 1963). For example, when *S*s wearing prisms are permitted to move their arms actively in the intertrial period, they adapt to the distortion far more rapidly and completely than *S*s who remain passive during this period (Singer & Day, 1966). An active *S* presumably obtains more proprioceptive, kinesthetic, and tactual information than a passive subject (Gibson, 1962). Until now, most investigations have involved the effects of sensory conflict on the ability to localize objects in space. The present experiments deal with form perception and examine whether the amount of information conveyed by each modality is also an important factor in this realm.

EXPERIMENT I

Rock and Victor (1964), in an experiment on form conflict, asked *S*s to hold an object through cloth while simultaneously viewing the object through a distorting element that changed its visual shape. When *S*s were asked to choose a match for the object from among other objects, they selected an object almost identical to the visual image. The question arises as to whether the visual dominance observed with a conflict task reflects a general relationship between vision and touch or whether a situation without conflict might not allow tactual information to make a greater contribution. In fact, using a task similar to that of Rock and Victor, but not involving conflict, Miller (1968) found that *S*s could match a tactual object without interference from the visual object. In this nonconflict task, *S*s viewed one object while at the same time feeling a different one. Thus, different visual and tactual stimulation was simultaneously available, but the information was not conflicting.

---

[1] This article is based on a PhD thesis submitted to the George Washington University. The research was supported in part by a National Science Foundation Institutional Traineeship and in part by National Institute of Mental Health Predoctoral Fellowship 1 F01 MH43167-01. My sincere thanks and esteem go to Richard D. Walk for his guidance through all phases of this research. I also thank Patricia Goldman for her excellent suggestions on the preparation of this paper and Jennie Lee Knight for preparing the figures.

[2] Requests for reprints should be sent to Epp Miller, Section on Neuropsychology, Laboratory of Psychology, National Institute of Mental Health, Room 1N-107, Building 9, Bethesda, Maryland 20014.

Three differences between the two tasks could account for the difference in the results. First, the presence or absence of conflict itself may be the critical factor. Second, *S*s in the conflict task were restricted to grasping the object passively, while in the nonconflict task *S*s were encouraged to feel the object actively. Third, in the conflict task *S*s were experimentally naive, but in the nonconflict task the procedures were explained beforehand and *S*s were presented with several pairs of objects.[3] As *S*'s knowledge about the task increases, through instructions and over trials, the likelihood may increase that he will change his approach to the task. Both haptic activity and cognitive changes may affect his attention and use of the available information (Canon, 1971). The results obtained with a conflict task and a nonconflict task may be more similar when haptic activity and task experience are equated with the two tasks. Experiment I, therefore, examined the effects of these two factors on matching performance using a form conflict task. Both active touch and the loss of procedural naiveté were expected to enhance the use of the tactual information and hence to produce matching closer to the true, tactual object.

Method

Subjects.—The *S*s were 64 college students, distributed equally among four groups.

Apparatus.—The stimulus objects were presented to *S* through the back of a box *S* held on his lap. (See Fig. 1.) The *S* looked into an open-ended cylinder suspended from the top of the box and that housed the distorting element. The *S* could thus view the stimulus object presented on a rod and pedestal directly under the end of the cylinder. The distorting element was a transparent Plexiglas rectangle (6.35 mm. thick) that had been bent to a curvature of about one-half a cylinder (Rock & Victor, 1964). The element compressed the image of the object in one dimension so that the shape of the viewed object was distorted. A tensor lamp inside the box eliminated the glare from the Plexiglas.

While *S* viewed the object, he could grasp or feel the stimulus object through side holes. A cotton cloth was placed over his hand so he could not see his fingers.

Stimulus objects.—Three aluminum objects were the stimuli: a rectangle, an oval, and a triangle. Each object's top surface defined its shape, and each stimulus object was 12.70 mm. thick. The rectangle and the oval were each 25.40 mm. high and 29.46 mm. wide. The triangle was 37.08 mm. high and had a base of 38.10 mm. The objects as seen through the distorting element were reduced by a little less than one-half in one dimension. The rectangular and the oval objects were reduced in width to 19.05 mm. The triangular object was reduced in height to 21.34 mm.

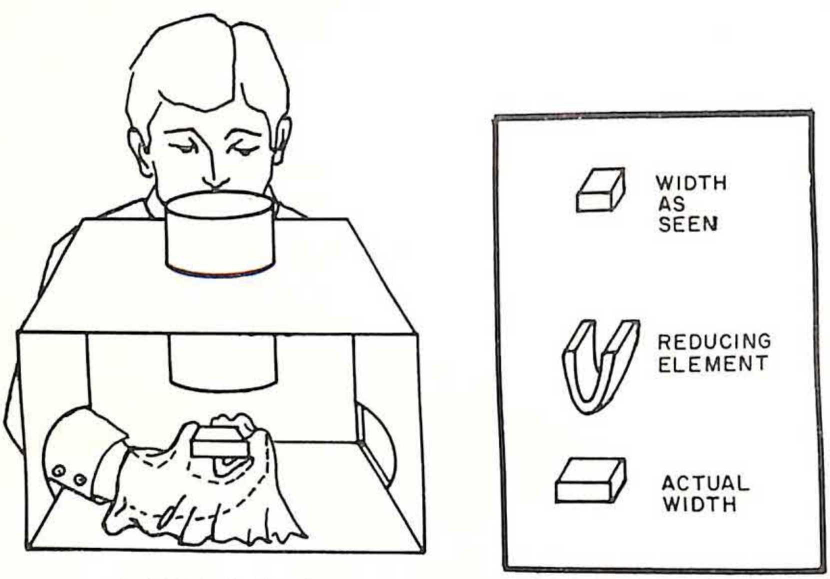
Response arrays and scoring.—Each stimulus object was one of a graduated series of 15 objects used as the response array. (See Fig. 2.) The oval and rectangular series were graduated in the width dimension only and in steps of 2.11 mm. The triangular series was graduated in the height dimension only and in steps of 2.54 mm. The objects were thus graduated in shape and were numbered from 3 to 17 to correspond with the number of steps present in the changing dimension. The number of the object chosen by *S* was the score given for that match. The actual objects used as the stimulus objects matched Object 14 on the array, and the visual appearance as distorted had a match of 9.1, as judged by an independent group of 16 *S*s. When matching the object, *S* could choose an object that would be the correct visual match, 9, the correct tactual match, 14, or any of the other objects. The objects were presented in random order for matching.

Design and procedures.—Each trial consisted of a presentation phase, when the object was presented for viewing and feeling, and a matching phase, when *S* was to pick out the object just presented. Matching was done by touch.[4] Although the primary interest was in the effects of active touch as opposed to passive handling of an object in the presentation phase, the conditions of matching themselves could influence the effects of stimulus presentation. To

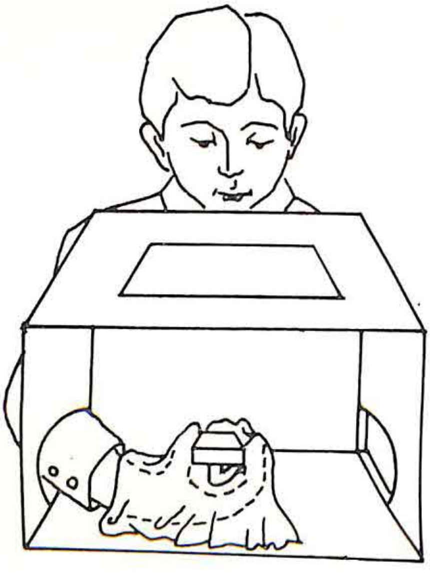
---

[3] Rock and Victor's (1964) *S*s who suspected deception or who used active touch responded differently than naive *S*s and passive *S*s and were eliminated from the sample. By contrast, the aim here was to study these factors.

[4] Typically with these procedures, similar results are obtained whether matching is done by vision or by touch (Miller, 1968, 1971; Rock, 1966; Rock & Victor, 1964).



EXPERIMENTAL

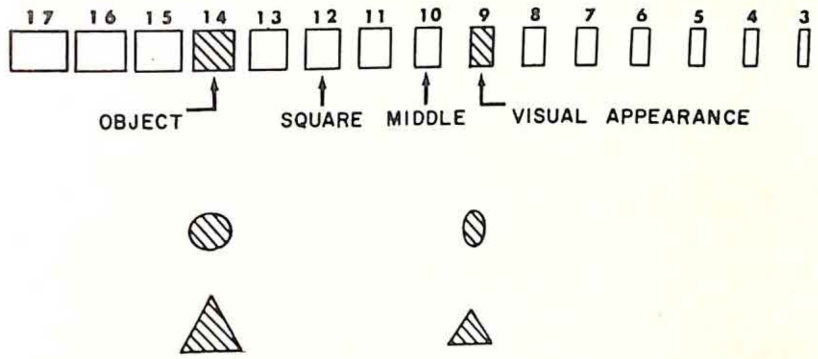


CONTROL

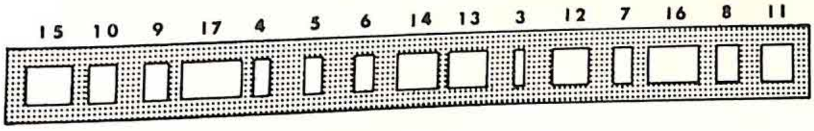
FIG. 1. Object presentation in the conflict task.

check on this possibility, some *S*s were instructed to use active touch, and some, passive touch, in each phase. Four combinations of active and passive touch in presentation and matching were possible, but the following three were compared: (*a*) Group 1, passive presentation followed by active matching (after Rock & Victor, 1964); (*b*) Group 2, active presentation, active matching; and (*c*) Group 3, passive presentation, passive matching.

When passive touch was used during the presentation phase, *S*s were limited to a single grasp of the object. In addition, when matching by passive touch, *S*s were presented with one choice object at a time and were limited to a single grasp of each.



THEORETICAL RESPONSE ARRAYS



ACTUAL RESPONSE ARRAY,OBJECTS IN RANDOM ORDER

FIG. 2. Response arrays.

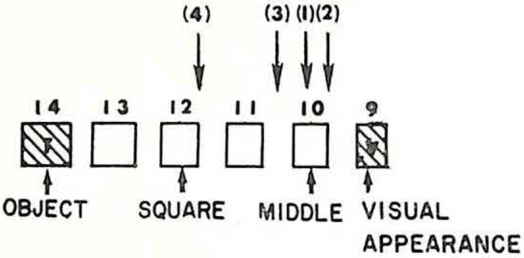
When active touch was used in the presentation phase, *S*s were instructed to grasp each object many times from different directions. When matching by active touch, *S*s were allowed free haptic exploration of the whole row of multiple-choice items.

In addition to these three groups given varied haptic activity in presentation and matching, Group 4, a control group, only felt the object during presentation and did not see the object at all. Group 4 used active touch both in presentation and in matching. Comparisons of Groups 1, 2, and 3 with Group 4 could be expected to show the influence of vision over touch. Comparisons among Groups 1, 2, and 3 should show the effect of haptic activity on visual dominance.

The effect of experience with the task was examined by studying each *S*'s performance over the first three trials. When the first object was presented, *S* was led through the procedure step by step. He was promised an explanation of the task when he was finished. The *S* was to look at an object and to grasp or feel it for 5 sec. (The *S*s in Group 4 only felt, and did not see, the object for 5 sec.) The *E* helped *S* find the object while uncovering the top of the viewing cylinder. The *E* observed whether *S* followed instructions to look and to feel simultaneously and to be appropriately active or passive in handling the object. (Two *S*s were rejected for feeling the object actively when they had been instructed to be passive.)

After 5 sec., *S* was asked to tie on a blindfold and then to choose the object from among "some very similar objects" placed on top of the box. For *S*s who matched using active touch, the 15 multiple-choice objects arranged in random order, as illustrated in Fig. 2, were placed on top of the box. The *S*'s hand was placed on either the left or right end, and he was asked to explore the objects and to choose the one in the box. The *S*s who used passive touch for matching (Group 3) were presented with the same 15 choice items, 1 at a time, and after a passive grasp of each, responded in turn as to whether or not the item was the object in the box. After the first match was completed, the second object was presented and matched, followed by the third object. Four orders of presenting the stimuli were balanced across *S*s. After the first trial, *S*s knew the task demands, and this loss of procedural naiveté was expected to lead to more veridical matches on the second and third trials.

A fourth trial was analyzed separately. The rectangular object was repeated, but it was presented for 30 sec. instead of 5 sec. Extending the exposure time was another attempt to decrease experimental naiveté by giving *S* more time to detect a discrepancy between vision and touch, and thereby to use the available tactual information to compensate for the distortion. The group that used active touch when the object was presented was expected to show the most compensation for the distortion.



GROUP	AVERAGE MATCHES
(1)	10.1
(2)	9.8
(3)	10.5
(4)	11.8

FIG. 3. Average matches of the conflict task treatment groups.

Last, a verbal questionnaire was administered to determine what each *S* expected and what he was aware of. The following questions were asked:

> [First] Did you think what you were seeing was not exactly as it felt in shape or size? Did you suspect any visual distortions? [Second] What did you experience? Look through the tube now and grasp the object without the cloth. Remember where your fingers were on the object. Take the object out of the box. Is that what you thought the object was like?

Results and Discussion

The purpose of this experiment was to reduce the exclusive reliance on vision in a situation like that of Rock and Victor (1964), in which the shape information available to vision and to touch was conflicting. Both haptic activity and the loss of procedural naiveté were expected to yield more veridical matching.

When vision and touch were in conflict, vision dominated touch, confirming the results of previous studies. The matches of the three conflict groups were close to the visual appearance of the object. (See Fig. 3.) A 4 × 3 repeated-measures analysis of variance of the first three trials revealed significant differences among the four groups, $F\ (3, 60) = 6.03, p < .01$. Each of the conflict conditions, Group 1, 2, and 3, was significantly different from the Touch Control group, Group 4 (Newman-Keuls test, $p < .01$), even though Group 4's matches were not completely accurate. Also, there were no significant differences among the three conflict groups that had used different combinations of active and passive touch. Thus, visual dominance was not diminished by haptic activity as predicted.

The loss of procedural naiveté also did not produce matches closer to the tactual object, as there were no significant differences among the first three trials, $F\ (2, 120) = .25$. The opportunity provided by Trials 2 and 3 was not sufficient to produce the cognitive and attentional changes expected to result in greater use of the tactual information. Even on Trial 4, when *S*s felt the object actively for 30 sec., they did not notice the discrepancy but continued to match the visual appearance of the object. After *S*s in the conflict groups were questioned about what they had experienced and what they had expected, they were shown the nature of the distortion. All of the *S*s were surprised, and no one reported an accurate awareness of the nature and amount of the distortion.

Visual dominance in this form conflict task was certainly more powerful than expected. From experiments on localization conflict, activity and task experience could reasonably be expected to decrease visual dominance to some degree. Since both haptic activity and task experience were ineffective, one possibility is that the resolution of a form discrepancy requires more tactual information than the resolution of a localization discrepancy. One could examine the question by extending the number and amount of conflict exposures with this task. Such a formulation would support the general amount-of-information hypothesis in sensory interaction but would require a different level or amount of information for form perception than for localization in space as suggested by many investigators (Gibson, 1970; Held, 1971; Pick & Hay, 1964; Rock, 1966).

While in this conflict task, the amount of tactual information appeared ineffective in changing visual dominance; in a situation without conflict, the amount of tactual information may yet be important in how vision and touch interact. Experiment II examined such a possibility.

Experiment II

Rock and Victor's (1964) form conflict task led to visual dominance over touch, and this finding was confirmed by Exp. I, despite the increase in haptic activity permitted as well as the loss of procedural naiveté. By contrast, on a form task that did not involve conflict, *S*s matched the tactual object independently of the visual information (Miller, 1968, 1971). Since the procedures in the nonconflict task employed active touch and fully instructed *S*s, these two factors may have eliminated the influence of vision by increasing the use of the tactual information. Perhaps if haptic activity and knowledge about task procedures were restricted, vision might influence touch even in this nonconflict situation. Such a result would be evidence that the reliance on vision or touch depends on the amount of information obtained by each. The purpose of Exp. II was, therefore, to explore active and passive touch and the loss of procedural naiveté using a form task that does not involve conflict.

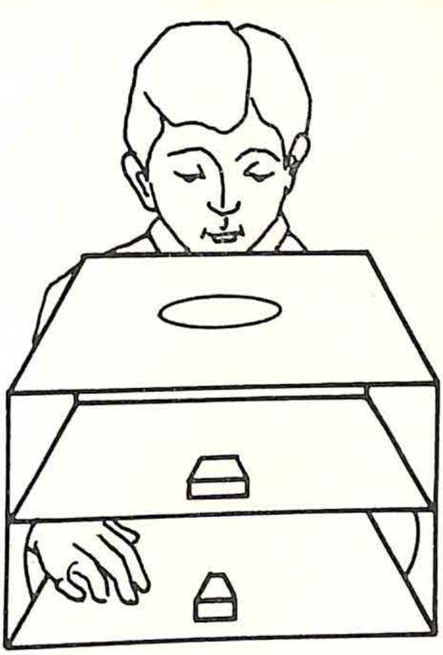
Method

Subjects.—The *S*s were 60 college students. Twenty *S*s were assigned to control conditions and 40 served as experimental *S*s.

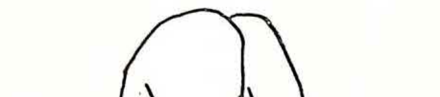
Apparatus.—The box used in Exp. I was modified in the following way. A shelf in the middle of the box was used for presentation of the visual stimuli viewed by the *S*s through a hole in the top. The tactual stimuli were presented on the bottom of the box.

Stimuli, response arrays, and scoring.—The materials described for the conflict task were used in this nonconflict task. The only difference was that the visual object in this task was 14 and the tactual object was 9. The arbitrary choice of these objects was made before Exp. I was conducted.

Design and procedures.—This experiment investigated the effects of active and passive touch and of the loss of procedural naiveté as Exp. I did, but the experimental design was somewhat different. Four factors were studied. First, there were two conditions of stimulus presentation. In the Experimental condition, one object was visually presented at the same time that a second and different object was tactually presented. In the Control condition, both objects were presented tactually, one to each hand. All of the *S*s were asked to match only the tactual object (the object felt by *S*'s preferred hand), permitting assessment of the degree to which one of the objects could influence information from the other object through vision (the Experimental condition) or through the same modality (the Control condition). For each condition, half of the *S*s were in-

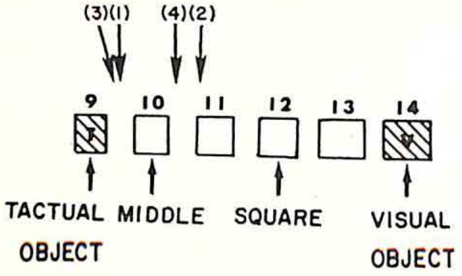


EXPERIMENTAL



CONTROL

Fig. 4. Object presentation in the nonconflict task.



PRESENTATION CONDITION	AVERAGE MATCH
(1) ACTIVE TOUCH, EXPERIMENTAL	9.6
(2) PASSIVE TOUCH, EXPERIMENTAL	10.9
(3) ACTIVE TOUCH, CONTROL	9.5
(4) PASSIVE TOUCH, CONTROL	10.4

FIG. 5. Nonconflict task results: Average matches.

structed to use active touch and half, passive touch. (Six *S*s were rejected for not following passive touch instructions.) During matching, all of the *S*s were encouraged to be active, unlike Exp. I, and a third factor, matching modality, was investigated instead. Half of the *S*s in each subgroup matched the tactual object by vision and half by touch. This factor was included in order to verify the results of previous work with this task. The fourth factor was the repeated-trials factor. Each *S* had three trials, one with the rectangular objects, one with the oval objects, and one with the triangular objects. The objects were presented in the constant order named.[5] As in Exp. I, *S*s were naive during the first presentation (they were led through the procedure step by step) and were most likely to be influenced by vision on this trial. After the first trial, *S*s knew the procedure and their performance was expected to improve.

Results and Discussion

Despite the simultaneity of the different visual and tactual stimulation in the experimental situation, vision did not dominate or influence touch. A 2 × 2 × 2 × 3 repeated-measures analysis of variance revealed no difference between the Experimental group (one object to vision and another to touch) and the Control group (two objects to touch), F (1, 52) = .49. The similarity of the tactual object matches in the Experimental and the Control groups is in striking contrast to the results of Exp. I in which the Experimental and Control groups were significantly different. When visual and tactual information was not only simultaneous but also conflicting as in Exp. I, vision completely dominated touch. When visual and tactual information was simultaneous but not conflicting as in the present experiment, touch was not influenced by vision. (See Fig. 5.)

Exploring the object actively produced more veridical matches than when passive touch was employed, but this variable was equally potent for the Experimental and the Control groups, F (1, 52) = 6.93, $p < .05$. Again, this result differs from that of Exp. I, in which no differential effects were observed with different amounts of haptic activity. Matching by vision was not significantly different from matching by touch, F (1, 52) = 2.16.

The loss of procedural naiveté did not produce the expected results. For the first trial, vision was expected to influence touch somewhat and this influence of vision was expected to decrease with subsequent trials. These results were not obtained. The tactual object was matched as accurately after the first presentation as after the second and the third, F (2, 104) = 3.03. While the first, second, and third matches were not significantly different from each other, two interactions with this factor were significant. One was an interaction between trials and stimulus presentation conditions, F (2, 104) = 6.38, $p < .01$. The Experimental *S*s matched just as accurately after the first presentation as they did after the second and third, but the Control *S*s matched more accurately after the first presentation than subsequently. This change over trials was unexpected both in its direction and by the groups that manifested it. The second interaction, F (2, 104) = 5.69, $p < .01$, was between trials and matching modality. Matching by vision or by touch gave similar results for the first two objects. For the third object, these tactual object matches were more toward the visual object when matching by touch and more toward the tactual object when matching by vision. Explanations for both of these interactions are not immediately apparent. In extensive work with this nonconflict task similar interactions did not appear, suggesting perhaps that they were fortuitous here.

---

[5] Additional research (Miller, 1971) compared three orders of presenting stimuli and indicated no significant differences among them. Hence, the constant order of presenting stimuli was used for convenience.

The present results, obtained on a nonconflict task, do not parallel those obtained with the conflict task. In the conflict task, vision dominated touch, despite haptic activity and continued experience with the task. In the nonconflict task, touch was essentially independent of vision even when the task experience was limited as on Trial 1, and passive touch was employed. Thus, increasing or decreasing the amount of tactual information does not seem to be an important determiner of how vision and touch interact in form perception. The contrast between the results obtained with these two tasks leads one to focus on still other differences between conflict and nonconflict situations. A third experiment was designed to explore some aspects of conflict itself.

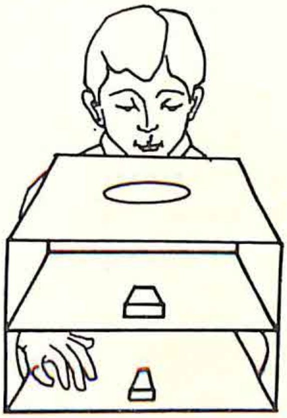
Experiment III

In a form conflict situation in which *S* is instructed to view and to feel an object, the conflict may occur because *S* expects this object to look and to feel the same. Further, since the object is localized in the same place both visually and tactually, the expectation of dealing with a single object may be confirmed. The purpose of the present experiment was to examine the roles of these two factors—the expectation and the localization experience—in inducing a sensory conflict and the visual dominance that characterizes it.

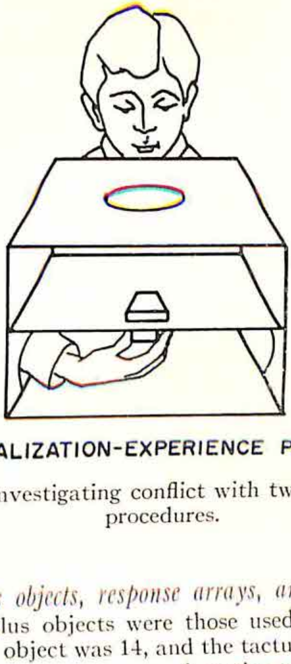
Method

Subjects.—The *Ss* were 40 college students, assigned randomly to one of two groups.

Apparatus.—The box used for the nonconflict task in Exp. II was used also by *Ss* who demonstrated the effect of the expectation of dealing with a single object. A variation of this box required by the Localization Experience group was introduced for this procedure. The visual object was presented on a shelf in the box as before. The tactual object, however, was presented on the underside of the shelf, directly under the visual object. (See Fig. 6.)



THE DECEPTIVE-INSTRUCTIONS PROCEDURE



THE LOCALIZATION-EXPERIENCE PROCEDURE

Fig. 6. Investigating conflict with two nonconflict procedures.

Stimulus objects, response arrays, and scoring.—The stimulus objects were those used in Exp. II. The visual object was 14, and the tactual object was 9. The response arrays and scoring were also like those in Exp. I and II.

Design and procedures.—Two groups, an Expectation group and a Localization Experience group, were each given a different nonconflict procedure. The effect of *expecting a single object* was investigated by giving deceptive instructions. The *Ss* were given one object to view and a different object to feel, but first they were told falsely they would be seeing and feeling identical halves of the same object. The *Ss* would see one half on the shelf and would feel the other half on the bottom of the box without seeing his hand. The *S* was promised an explanation of this procedure later. Then *S* was to match the object. If *S* asked whether the objects were the same or not during the experiment, *E* re-

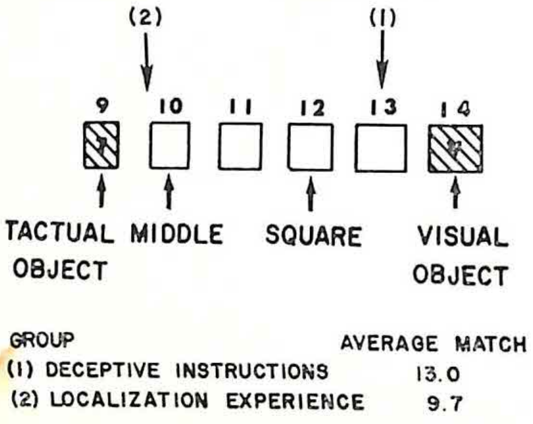


FIG. 7. Average matches of Exp. III's treatment groups.

plied that they were identical halves of the same object. Actually, the visual object was 14, very different from the tactual object, 9.

For the Localization Experience group *E* placed two different objects back-to-back inside the box, so they had essentially the same spatial location. The *S*s in this group were to view the visual object on the shelf while feeling the tactual object on the underside of the shelf. The actual instructions were ambiguous. The *S*s were instructed to look at "something" and feel "something" and then to match "the object." These instructions would be appropriate whether *S*s experienced the two objects as being the same or different. One could determine, then, what *S*s' uninstructed localization experience of the objects would be and how it would correspond with matching performance. If *S*s thought the objects were different, they would match on the basis of one of them, as with the other nonconflict task; but the common spatial location of the objects might still lead to visual influence over touch.

For convenience, *S*s in both of these groups matched the objects by vision. The second factor investigated the effects of both haptic activity and trials. Each *S* was given three trials, one each with the rectangles, the ovals, and the triangles. On the first trial, he was led through the procedure step by step. In addition, each *S* was limited to a passive grasp for the first two trials. On the third trial, *S* was instructed to feel the tactual object actively. Thus, the difference between the first and second trials would suggest the effect of the loss of procedural naiveté, and the difference between the second and third trials indicates the effects of passive and active touch as well. (Five *S*s were rejected because they did not follow passive touch instructions.)

When the presentations had been given, *E* asked all *S*s three questions to determine further how each *S*'s expectations may have influenced his matches. First, they were asked whether the two objects of each pair were always identical to each other, always different, or the same for some object pairs and different for others. Second, *S*s were asked if they initially expected the objects to be the same, different, or were open-minded about what to expect. Then they were asked whether they relied more on vision or on touch. When asked pointedly, no *S*s replied that they had tried purposely to ignore one of the objects.

Results and Discussion

The Deceptive Instructions group, *S*s who were told that the objects were identical halves of the same object, had an average match of 13.0, close to the correct visual object match of 14. By contrast, the Localization Experience group's average match was 9.7, very close to the correct tactual object match of 9. A 2 × 3 repeated-measures analysis of variance revealed that the Deceptive Instructions group had significantly different matches from the Localization Experience group, $F\ (1, 38) = 36.02, p < .001$. (See Fig. 7.)

The results of the verbal questionnaire supported this difference between the groups. The *S*s in the Deceptive Instructions group reported they had been led to expect the two objects of each pair to be identical halves of the same object. Most of the *S*s then perceived the objects as the same and matched relying on vision. The expectation or assumption that a single object was involved led to visual dominance similar to that observed in the conflict task. On the other hand, the Localization Experience *S*s reported that they had been led to expect the object of a pair to be different or were open-minded. They perceived the objects as different and matched the tactual object. The experience of a common location for the different visual and tactual stimuli, an experience that is part of a form conflict situation, led to no more use of vision than the other nonconflict task had.

The loss of experimental naiveté appeared to have no effect, as no difference was obtained between the first trial and the second trial. The differences between passive touch and active touch were not seen in this experiment. The second trial, using passive touch, was as accurate as the third trial, performed with active touch. In fact, there were no significant differences among the three trials, $F\ (2, 76) = 1.14$.

The questionnaire also confirmed that *S*s did not change their task approach over trials.

The localization experience procedure yielded accurate tactual object matches, independent of vision, as had other nonconflict procedures. Merely the same localization of two sources of stimulation by vision and by touch does not produce visual dominance. This localization experience alone, then, seems not to be an important aspect of a conflict task. By contrast, the deceptive instructions procedure led to reliance on vision. When *S* believed a single object was involved, vision appeared to be dominant, as in a conflict task. This finding emphasizes a more cognitive component involved in how the perceptual modalities work together. That is, what one expects and what one already knows about a situation appear to influence how the information available is handled.

REFERENCES

CANON, L. K. Directed attention and maladaptive "adaptation" to displacement of the visual field. *Journal of Experimental Psychology*, 1971, **88**, 403–408.

FREEDMAN, S. J. On the mechanisms of perceptual compensation. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior*. Homewood, Illinois: Dorsey Press, 1968.

GIBSON, E. J. The development of perception as an adaptive process. *American Scientist*, 1970, **58**, 98–107.

GIBSON, J. J. Observations on active touch. *Psychological Review*, 1962, **69**, 477–491.

HAY, S. C., PICK, H. L., & IKEDA, K. Visual capture produced by prism spectacles. *Psychonomic Science*, 1965, **2**, 215–216.

HELD, R. Plasticity in sensory-motor systems. *Scientific American*, 1965, **213**, 84–94.

HELD, R. Two modes of processing spatially distributed visual stimulation. In F. O. Schmitt (Ed.), *The neurosciences: Second study program*. New York: Rockefeller University Press, 1971.

HOWARD, I. P., CRASKE, B., & TEMPLETON, W. B. Visuo-motor adaptation to discordant exafferent stimulation. *Journal of Experimental Psychology*, 1965, **70**, 189–191.

MILLER, E. A. The interaction of vision and touch in a matching task. Paper presented at the meeting of the Eastern Psychological Association, Washington, D.C., April 1968.

MILLER, E. A. The interaction of vision and touch in form perception with a conflict task and a nonconflict task. Unpublished doctoral dissertation, George Washington University, 1971.

MOULDEN, B. Adaptation to displaced vision: Reafference is a special case of the cue-discrepancy hypothesis. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 113–117.

OVER, R. An experimentally induced conflict between vision and proprioception. *British Journal of Psychology*, 1966, **57**, 335–341.

PICK, H. L., & HAY, J. C. Adaption to prismatic distortion. *Psychonomic Science*, 1964, **1**, 199–200.

ROCK, I. *The nature of perceptual adaptation*. New York: Basic Books, 1966.

ROCK, I., & HARRIS, C. S. Vision and touch. *Scientific American*, 1967, **216**, 96–104.

ROCK, I., & VICTOR, J. Vision and touch: An experimentally created conflict between the two senses. *Science*, 1964, **143**, 594–596.

SINGER, G., & DAY, R. H. Spatialad aptation and after effects with optically transformed vision: Effects of active and passive responding and the relationship between test and exposure responses. *Journal of Experimental Psychology*, 1966, **71**, 725–731.

WALLACH, H., & KARSH, E. B. The modification of stereoscopic depth-perception and the kinetic depth-effect. *American Journal of Psychology*, 1963, **76**, 429–435.

(Received February 4, 1972)

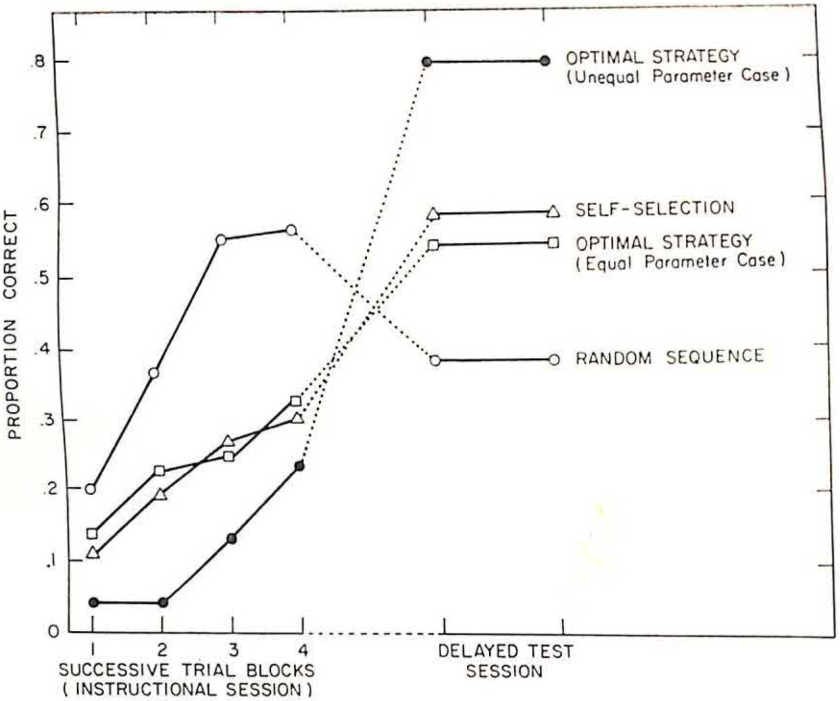


FIG. 1. Proportion of correct responses in successive trial blocks during the instructional session and on the delayed test administered 1 wk. later.

figure, data are presented for performance during the instructional session; on the right side are results from the delayed test. The data from the instructional session are presented in four successive blocks of 84 trials each; for the RO condition this means that on the average each item was presented once in each of these blocks. Note that performance during the instructional session is best for the RO condition, next best for the OE condition which is slightly better than the SS condition, and poorest for the OU condition; these differences are highly significant, F (3, 116) = 21.3, $p < .001$. The order of the experimental groups on the delayed test is completely reversed. The OU condition is by far best with a correct response probability of .79; the SS condition is next with .58; the OE condition follows closely at .54; and the RO condition is poorest at .38, F (3, 116) = 18.4, $p < .001$. The observed pattern of results is what one would expect. In the SS condition, *S*s are trying to test themselves on items they do not know; consequently, during the instructional session, they should have a lower proportion of correct responses than *S*s run on the RO procedure where items are tested at random. Similarly, the OE and OU conditions involve a procedure that attempts to identify and test those items that have not yet been mastered and also should produce high error rates during the instructional session. The ordering of groups on the delayed test is reversed since now all words are tested in a nonselective fashion; under these conditions the proportion of correct responses provides a measure of *S*'s mastery of the total set of vocabulary items.

The magnitude of the effects observed on the delayed test are large and of practical significance. The SS condition (when compared to the RO condition) leads to a relative gain of 53%, whereas the OU condition yields a relative gain of 108%. It is interesting that *S* can be very effective in determining an optimal study sequence, but not so effective as the best of the two adaptive teaching systems.

DISCUSSION

At this point, we turn to an account of the theory on which the OU and OE schemes are based. Both schemes assume that acquisition of a second-language vocabulary can be described by a fairly simple learning model. It is postulated that a given item is in one of three states (P, T, and U) at any moment in time. If the item is in State P, then its translation is known and this knowledge is "relatively" permanent in the sense that the learning of other items will not interfere with it. If the item is in State T, then it is also known but on a "temporary" basis; in State T the learning of other items can give rise to interference effects that cause the item to be forgotten. In State U the item is not known, and *S* is unable to give a translation. Thus in States P and T a correct translation is given with probability one, whereas in State U the probability is zero.

When Item *i* is presented for test and study the following transition matrix describes the possible change in state from the onset of the trial to its termination:

$$\mathbf{A}_i = \begin{array}{c} \\ P \\ T \\ U \end{array} \begin{array}{c} \begin{array}{ccc} P & T & U \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ x_i & 1 - x_i & 0 \\ y_i & z_i & 1 - y_i - z_i \end{bmatrix} \end{array}.$$

Rows of the matrix represent the state of Item *i* at the start of the trial and columns the state at the end of the trial. On a trial when some item *other* than Item *i* is presented for test and study (whether that item is a member of Item *i*'s display list or some other display list), transitions in the learning state of Item *i* also may take place. Such transitions can occur only if *S* makes an error on the trial; in that case the transition matrix applied to Item *i* is as follows:

$$\mathbf{F}_i = \begin{array}{c} \\ P \\ T \\ U \end{array} \begin{array}{c} \begin{array}{ccc} P & T & U \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 - f_i & f_i \\ 0 & 0 & 1 \end{bmatrix} \end{array}.$$

Basically, the idea is that when some other item is presented to which *S* makes an error (i.e., an item in State U), then forgetting may occur for Item *i* if it is in State T.

To summarize, when Item *i* is presented for test and study transition Matrix $\mathbf{A}_i$ is applied; when some other item is presented that elicits an error then Matrix $\mathbf{F}_i$ is applied. The above assumptions provide a complete account of the learning process. For the task considered in this article it is also assumed that Item *i* is either in State P (with probability g_i) or in State U (with probability $1 - g_i$) at the start of the instructional session; *S* either knows the correct translation without having studied the item or does not. The parameter vector $\phi_i = [x_i, y_i, z_i, f_i, g_i]$ characterizes the learning of a given Item *i* in the vocabulary set. The first three parameters govern the acquisition process; the next parameter, forgetting; and the last, *S*'s knowledge prior to entering the experiment.

For a more detailed account of the model the reader is referred to Atkinson and Crothers (1964) and Calfee and Atkinson (1965). It has been shown in a series of experiments that the model provides a fairly good account of vocabulary learning and for this reason it was selected to develop an optimal procedure for controlling instruction. We now turn to a discussion of how the OE and OU procedures were derived from the model. Prior to conducting the experiment reported in this article, a pilot study was run using the same word lists and the RO procedure described above. Data from the pilot study were employed to estimate the parameters of the model; the estimates were obtained using the minimum chi-square procedures described in Atkinson and Crothers. Two separate estimates of parameters were made. In one case it was assumed that the items were equally difficult, and data from all 84 items were lumped together to obtain a single estimate of the parameter vector ϕ: this estimation procedure will be called the equal-parameter case (E case), since all items are assumed to be of equal difficulty. In the second case data were separated by items, and an estimate of ϕ_i was made for each of the 84 items (i.e., $84 \times 5 = 420$ parameters were estimated); this procedure will be called the unequal-parameter case (U case). In both the U and E cases, it was assumed that there were no differences among *Ss*; this homogeneity assumption regarding learners will be commented upon later. The two sets of parameter estimates were used to generate the optimization schemes previously referred to as the OE and OU procedures; the former based on estimates from Case E and the latter from Case U.

In order to formulate an instructional strategy, it is necessary to be precise about the quantity to be maximized. For the present experiment the goal is to maximize the total number of items *S* correctly translates on the delayed test.[2] To do this, we need to specify the theoretical relationship between the state of learning at the end of the instructional session and performance on the delayed test. The assumption made here is that only those items in State P at the end of the instructional session will be translated correctly on the delayed test; an item in State T is presumed to be forgotten during the intervening week. Thus, the problem of maximizing delayed-test performance involves, at least in theory, maximizing the number of items in State P at the termination of the instructional session.

Having numerical values for parameters and knowing *S*'s response history, it is possible to estimate his current state of learning.[3] Stated more precisely, the learning model can be used to derive equations and, in turn, compute the probabilities of being in States P, T, and U for each item at the start of Trial n, conditionalized on *S*'s response history up to and including Trial $n - 1$. Given numerical estimates of these probabilities, a strategy for optimizing performance is to select that item for presentation (from the current display list) that has the greatest probability of moving into State P if it is tested and studied on the trial. This strategy has been termed the one-stage optimization procedure because it looks ahead only one trial in making decisions. The true optimal policy (i.e., an N-stage procedure) would consider all possible item-response sequences for the remaining trials and select the next item so as to maximize the number of items in State P at the termination of the instructional session. Unfortunately, for the present case the N-stage policy cannot be applied because the necessary computations are too time consuming even for a large-scale computer. A series of Monte Carlo studies indicates that the one-stage policy is a good approximation to the optimal strategy for a variety of Markov learning models; it was for this reason, as well as the relative ease of computing, that the one-stage procedure was employed. For a discussion of one-stage and N-stage policies and Monte Carlo studies comparing them see Groen and Atkinson (1966), Calfee (1970), and Laubsch (1970).

The optimization procedure described above was implemented on the computer and permitted decisions to be made on-line for each *S* on a trial-by-trial basis. For *S*s in the OE group, the computations were carried out using the five parameter values estimated under the assumption of homogeneous items (E-case); for *S*s in the OU group the computations were based on the 420 parameter values estimated under the assumption of heterogeneous items (U-case).

The OU procedure is sensitive to interitem differences and consequently generates a more effective optimization strategy than the OE procedure. The OE procedure, however, is almost as effective as having *S* make his own instructional decisions and far superior to a random presentation scheme. If individual differences among *S*s also are taken into account, then further improvements in delayed-test performance should be possible; this issue and methods for dealing with individual differences are discussed in Atkinson and Paulson (1972).

The study reported here illustrates one approach that can contribute to the development of a theory of instruction (Hilgard, 1964). This is not to suggest that the OU procedure represents a final solution to the problem of optimal item selection. The model upon which this strategy is based ignores several important factors, such as interitem relationships, motivation, and short-term memory effects (Atkinson & Shiffrin, 1968, p. 190). Undoubtedly, strategies based on learning models that take these variables into account would yield superior procedures.

Although the task considered in this article deals with a limited form of instruction, there are at least two practical reasons for studying it. First, this type of task occurs in numerous learning situations; no matter what the pedagogical orientation, any initial reading program or foreign-language course involves some form of list learning. In this regard it should be noted that a modified version of the OU strategy has been used successfully in the Stanford computer-assisted instruction program in initial reading (Atkinson & Fletcher, 1972). Second, the study of simple tasks that can be understood in detail provides prototypes for analyzing more complex optimization problems. At present, analyses comparable to those reported here cannot be made for many instructional procedures of central interest to educators, but examples of this sort help to clarify the steps involved in devising and testing optimal strategies.[4]

---

[2] Other measures can be used to assess the benefits of an instructional strategy; e.g., in this case weights could be assigned to items measuring their relative importance. Also costs may be associated with the various actions taken during an instructional session. Thus, for the general case, the optimization problem involves assessing costs and benefits and finding a strategy that maximizes an appropriate function defined on them. For a discussion of this issue see Atkinson and Paulson (1972), Dear, Silberman, Estavan, and Atkinson (1967), and Smallwood (1971).

[3] The *S*'s *response history* is a record for each trial of the vocabulary item presented and the response that occurred. It can be shown that there exists a *sufficient history* that contains only the information necessary to estimate *S*'s current state of learning; the sufficient history is always a function of the complete history and the assumed learning model (Groen & Atkinson, 1966). For the model considered in this paper the sufficient history is fairly simple. It is specified in terms of individual vocabulary items for each *S*; we need to know the ordered sequence of correct and incorrect responses to a given item plus the number of errors (to other items) that intervene between each presentation of the item.

---

[4] For a review of work on optimizing learning and references to the literature see Atkinson (1972).

REFERENCES

Atkinson, R. C. Ingredients for a theory of instruction. *American Psychologist,* 1972, **27,** 921–931.

Atkinson, R. C., & Crothers, E. J. A comparison of paired-associate learning models having different acquisitions and retention axioms. *Journal of Mathematical Psychology,* 1964, **2,** 285–315.

Atkinson, R. C., & Fletcher, J. D. Teaching children to read with a computer. *The Reading Teacher,* 1972, **25,** 319–327.

Atkinson, R. C., & Paulson, J. A. An approach to the psychology of instruction. *Psychological Bulletin,* 1972, **78,** 49–61.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 2. New York: Academic Press, 1968.

Calfee, R. C. The role of mathematical models in optimizing instruction. *Scientia: Revue Internationale de Synthèse Scientifique,* 1970, **105,** 1–25.

Calfee, R. C., & Atkinson, R. C. Paired-associate models and the effect of list length. *Journal of Mathematical Psychology,* 1965, **2,** 255–267.

Dear, R. E., Silberman, H. F., Estavan, D. P., & Atkinson, R. C. An optimal strategy for the presentation of paired-associate items. *Behavioral Science,* 1967, **12,** 1–13.

Groen, G. J., & Atkinson, R. C. Models for optimizing the learning process. *Psychological Bulletin,* 1966, **66,** 309–320.

Hilgard, E. R. (Ed.) *Theories of learning and instruction: The sixty-third yearbook of the National Society for the Study of Education.* Chicago: University of Chicago Press, 1964.

Laubsch, J. H. Optimal item allocation in computer-assisted instruction. *IAG Journal,* 1970, **3,** 295–311.

Smallwood, R. D. The analysis of economic teaching strategies for a simple learning model. *Journal of Mathematical Psychology,* 1971, **8,** 285–301.

(Received February 4, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 130–133

THE GENERALITY OF FITTS'S LAW

LISANNE BAINBRIDGE[1] AND MOIRA SANDERS

University of Reading, Reading, England

In P. M. Fitts's original equation describing the relation between amplitude, accuracy, and duration of hand movements, the duration depends on the logarithm of the ratio of amplitude and accuracy. Data presented by A. T. Welford suggest that the logarithmic transform gives additivity of amplitude and accuracy effects. This paper presents data from a variant of Fitts's dotting task, in which amplitude and duration of movements were controlled and accuracy measured. The results support the original ratio equation.

Fitts (1954) was the first to suggest an equation which related the amplitude, accuracy, and duration of hand movements. This equation was determined from empirical studies in which *S* moved his hand to and fro between two targets. The distance (A) between, and width (W) of, the targets was varied and the time (MT) taken to make the movements was measured. The following equation fitted the data:

$$MT = a \cdot \log_2(2A/W) + b,$$

in which a and b are constants.

From his equation, Fitts (1954) made inferences about the mechanisms of movement control in terms of information theory. Use of binary logarithms is not necessary to the analysis; however, the data would give a straight line if plotted using the above equation with logarithms to any base. Change of base, e.g., to $\log_x$, simply changes the constant a by a factor $\log_x(2)$. The multiplier, 2, is not an empirically determined parameter but was added by Fitts to ensure that the ratio was always positive; it has the effect of adding a constant. The essential aspects of Fitts's equation are that the movement time is a function of the ratio of the amplitude and accuracy of the movement, whatever their absolute sizes, and that this function is logarithmic.

Later researchers have studied the same type of task and suggested refinements of this equation. In Fitts's original equation, the A and W values were those defined by *E* and drawn as the targets. Since *S* may not be working according to the values thus defined, *E* should calculate the A/W ratios from *S*'s A and W values, A' and W', in which A' is the measured distance between the midpoints of the two hit distributions and W' is the measured width of the scatter of *S*'s hit points about a given target. For the present purpose, the most important refinements of Fitts's equation using A' and W' were given by Welford (1968, pp. 147 & 156):

$$MT = k \cdot \log(A'/W' + .5) \qquad [1]$$

$$MT = k \cdot \log(A'/(W' - c) + .5). \qquad [2]$$

In Welford's study, a single line could be drawn through the data on a plot of movement time against Equation 1, but a better fit to a line was obtained by plotting against Equation 2, with $c = 3$ mm. In a study by Welford (1969) in which the parameters were empirically determined, average values obtained were $k = .116$ and $c = 3.16$ mm., for *S*s aged 20–60 yr.

Welford (1968, p. 157) also presented a different analysis of the data, which brings into question the basic ratio relation, suggesting instead that the absolute sizes of A and W do affect the movement time. In Welford's plot of movement time against Equation 1, the points for one target width and different movement amplitudes did fall on a straight line. A set of such lines was obtained, one for each target width used. These lines were parallel. This means that a change in target width added a small constant to the movement time, and this constant was unaffected by the particular amplitude being tested. Parallel lines were also obtained if points for any one amplitude at different target widths were connected. Thus the effects of accuracy and amplitude could be analyzed separately by taking their logarithms. Their effects are then additive and their interaction does not have to be considered. The equation describing this relation is given by Welford (1968):

---

[1] Requests for reprints should be sent to Lisanne Bainbridge, Department of Psychology, University of Reading, Building 3, Earley Gate, Whiteknights, Reading, RG 6 2AL, England.

$$MT = a \cdot \log A' + b \cdot \log(1/W') \quad [3]$$

From Welford's data the ratio $b:a$, for *S*s aged 20–60 yr., was 1.75:1; i.e., in determining movement time accuracy was weighted 1.75 times more heavily than amplitude.

If $a = b$, Equation 3 simplifies to

$$\begin{aligned} MT &= a[\log A' + \log(1/W')] \\ &= a \cdot \log(A'/W'), \end{aligned}$$

the ratio as found in previous studies. This suggests that Fitts's finding, that movement time is determined by the ratio of amplitude and accuracy rather than their absolute values, may have been a happy accident of using a task in which amplitude and accuracy are equally weighted, rather than being a basic determinant of movement control.

Welford's (1969) analysis did not resolve which of these two formulations, the ratio or the additive, best fitted the data. The data plotted against Equation 2 with $c = 3$ fitted a straight line. The data plotted against Equation 3, with $a = .104$ and $b = .177$, fitted a straight line just as well. If one of these equations is a general statement of the relation between amplitude, accuracy, and duration, then it should apply in other tasks. This paper presents a study of a simple inversion of Fitts's original task. Amplitude and duration of movement were controlled, and accuracy measured. Brown and Slater-Hammel (1949) varied the amplitude of movements made to a single line target, and found that duration of movement increased with amplitude, at an essentially constant accuracy of within less than 2 mm. from the target line. Woodworth (1899) studied movements of constant duration made over amplitudes of 50, 100, 150, and 200 mm. The error scores were slightly less than would be predicted from Fitts's equation. In the experiment described here both amplitude and duration of movement were varied systematically. Analysis of the results should show whether the same type of relation generalizes to the description of this task, and if so should help to discriminate between the variant equations suggested for the original task.

TABLE 1

SUBJECTS' MOVEMENT ACCURACY (W') AND AMPLITUDE (A') FOR EACH CONDITION

MT (in sec.)	Item	A (in in.) 4	8	16
.3	W'	.99	1.75	3.13
	A'	4.06	7.77	14.66
.4	W'	.68	1.09	2.0
	A'	3.99	8.01	15.6
.6	W'	.34	.41	.79
	A'	3.99	8.01	15.86

Note.—Movement accuracy and amplitude are measured in inches.

METHOD AND DESIGN

The *S* dotted between lines drawn on paper. Each target was marked by a single line, so a target sheet contained a pair of parallel lines. The distance between the lines was the amplitude as defined by *E*. Three amplitudes were used, 4″ (102 mm.), 8″ (204 mm.), and 16″ (408 mm.). Movement time was set by making *S* move in time with the beats of a metronome. Three metronome speeds were used, 100, 150, and 200 movements per minute, or .6, .4 and .3 sec. per movement. Each trial lasted for 1 min. The *S* was asked to be as accurate as possible in hitting the target line, but the importance of keeping up with the metronome was stressed. The *S* was told that accuracy should suffer rather than timing if the beats were too fast.

The three amplitudes and three speeds gave nine experimental conditions. These nine conditions were presented in pseudorandom order to *S*. They were then presented again in the reverse order. Each *S* was given a different random order. Ten *S*s, all volunteer university students, were used.

The accuracy of hits around each target was found by the simple technique used by Welford (1969). W' was taken as the distance between extreme shots, measured along the axis of the movements, ignoring wild deviations. To measure *S*'s amplitude, A', a

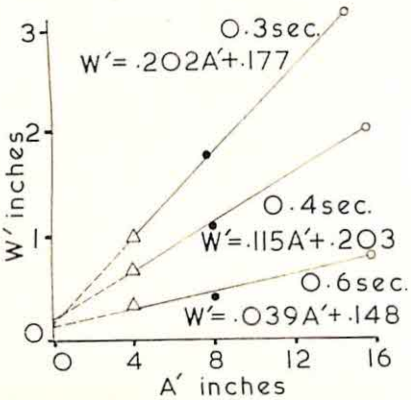


FIG. 1. Accuracy as a function of amplitude, with movement time as a parameter.

line was drawn down the linear center of the distribution of hits for each target, perpendicular to the axis of movement. The A' in a given trial was measured as the distance between these lines drawn for the two targets.

RESULTS

There were 20 readings for each condition; the means of these were calculated, in inches, and are given in Table 1.

Figure 1 shows the accuracy achieved with different movement amplitudes at the three levels of movement time. Inaccuracy (width of target) was linearly related to amplitude. This relation interacted with MT, as the slope of the function varied with MT. The equations for the three lines, obtained from least-squares regression calculations, are given in Fig. 1. The lines had a common intercept, of average value .176. The differences in slope were statistically significant. For the slopes at .6- and .4-sec. MT, $F\ (1, 2) = 40.34$; and for .4- and .3-sec. MT, $F\ (1, 2) = 44.95$—these were significant at $p < .05$; for the slopes at .6 and .3 sec., $F\ (1, 2) = 221.05, p < .01$. The relation of slope of line to MT could not be fitted by a linear, logarithmic, or power equation, but could be fitted by an exponential equation. In this equation, 2 is used as a base, as is common in these studies. Substituting this equation for slope in the general equation for these straight lines gives:

$$W' = (1.015 \times 2^{-7.8175MT})A' + .176. \quad [4]$$

For comparison with earlier equations we wish to find MT in terms of W' and A'. Taking logarithms of Equation 4 and solving for MT we obtain:

$$MT = .1279 \log_2(A'/(W' - .176)) + .0215. \quad [5]$$

Equation 5 is similar in form to Equation 2.

The target width W' was the dependent variable in this experiment. A version of Welford's (1968) additive type of equation, 3, in terms of W' would be:

$$\log W' = (1/b)(a \cdot \log A' - MT).$$

The graph of $\log_2 W'$ against $(\log_2 A' - MT)$ is shown in Fig. 2, which also shows the equations for the three lines of constant A'. The differences in slope of these lines were statistically significant. For the slopes at 4″ and 8″ A, $F\ (1, 2) = 40.36$; and for 8″ and 16″, $F\ (1, 2) = 88.48$, both significant at $p < .05$; for the slopes at 4″ and 16″, $F\ (1, 2) = 162.5$, $p < .01$. The slopes and intercepts of these lines both changed consistently with A', and both these functions were best fitted (by regression) by logarithmic equations. These equations substituted in the general linear equation give:

$$\log_2 W' = (2.71 \log_2 A' - .47)(\log_2 A' - MT) + (19.43 - 13.61 \log_2 A'). \quad [6]$$

This equation cannot be solved for MT in a way in which $\log W'$ and $\log A'$ are independent additive components.

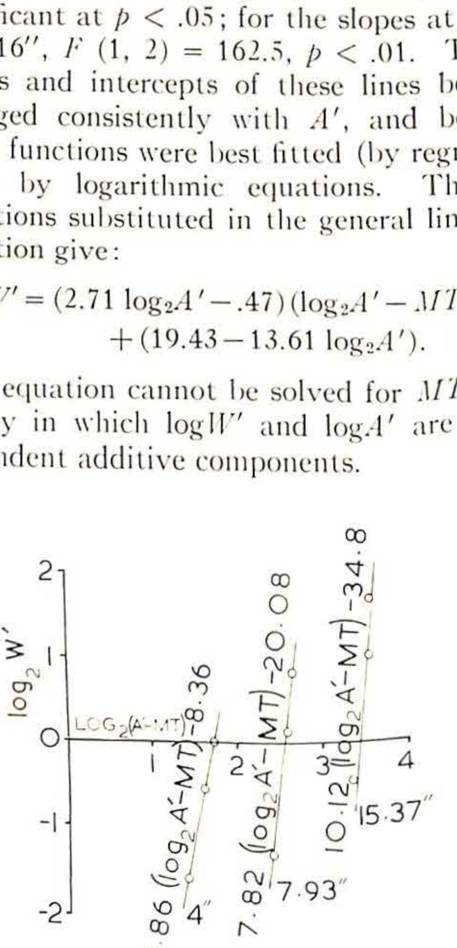


FIG. 2. $\text{Log}_2 W'$ as a function of $(\log_2 A' - MT)$, with mean amplitude A' as a parameter.

Discussion

We wish to test whether the data from this experiment is best fitted by an equation such as Equation 2, in which A' and W' interact as first suggested by Fitts (1954), or whether the logarithmic transform has the effect that A' and W' are independent and additive as suggested by Welford (1968) in Equation 3.

Two analyses of the data have been presented. Analysis of the untransformed data gave an equation, 5, which is very similar to Equation 2. In Equation 5, the constant .1279 is similar to Welford's k value of .116. Welford (1969) obtained a c calue of 3.16 mm. The constant .176 in Equation 5 can be taken as a measure in inches as used in this task and is equivalent to 4.47 mm. The discrepancy between these c values may be because c is a parameter which differs in the two tasks. Alternatively, Welford used a fixed added constant of .5 in his calculations, which may have influenced the c values he found. The data from this analysis therefore provides good support for an equation in which the effects of amplitude and accuracy interact, as suggested by Fitts.

Support for an equation like Equation 3 would be obtained from the data after log transform. If the log transform imposed additivity of A' and W', the lines in Fig. 2 should be parallel; an increase in $\log A'$ should not affect the slope of the line relating $\log W'$ and MT. As shown in this article, the slopes of these lines are not the same, and they vary consistently with A'. Also, it should be possible to solve for MT the one equation derived for all the lines in such a way that $\log W'$ and $\log A'$ are independent additive components as in Equation 3, and this cannot be done for Equation 6. The data from this experiment therefore does not fit an independent additive equation like Equation 3, the type suggested by Welford (1968).

The results of an experiment on the amplitude, duration, and accuracy of movement, using a variant of the original Fitts dotting task, therefore further support Fitts's analysis that it is the ratio of amplitude and accuracy which determine the duration of a movement, and show that this relation describes performance in an inverse type of task.

REFERENCES

Brown, J. S., & Slater-Hammel, A. T. Discrete movements in the horizontal plane as a function of their length and duration. *Journal of Experimental Psychology*, 1949, **39**, 84–95.

Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 1954, **47**, 381–391.

Welford, A. T. *Fundamentals of skill*. London: Methuen, 1968.

Welford, A. T. Speed and accuracy of movement and their changes with age. *Acta Psychologica*, 1969, **30**, 3–15.

Woodworth, R. S. The accuracy of voluntary movements. *Psychological Review*, 1899, **3**(2, Whole No. 13).

(Received February 8, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 134–137

FACTORS IN MOTOR SHORT-TERM MEMORY: THE INTERFERENCE EFFECT OF INTERPOLATED ACTIVITY

ERIC A. ROY AND WILLIAM G. DAVENPORT [1]

University of British Columbia, Vancouver, British Columbia, Canada

The effect of number of interpolated trials on the short-term retention of a positioning response was investigated. Absolute and algebraic errors were the dependent measures. Three interpolated trial conditions, zero interpolated learning trials (0IL), two IL(2IL), four IL(4IL); and two retention intervals, 60 and 90 sec., were used. Significant interference occurred only in the 4IL condition and increased over the retention intervals. Results supported previous research showing that interference rather than trace decay was the major factor in forgetting of a discrete motor response. Similarities in forgetting patterns for verbal and motor short-term memory were discussed as well as recommendations for future studies.

In verbal short-term memory (STM), there is considerable support for interference as a cause of forgetting (Keppel & Underwood, 1962; Waugh & Norman, 1965). In motor STM, prior activity causing proactive interference (PI) has been implicated as a major source of forgetting (Ascoli & Schmidt, 1969; Stelmach, 1969; Williams, 1971). However, evidence of trace decay was present (Ascoli & Schmidt, 1969; Stelmach, 1969). A similar interference effect of interpolated activity has not been as extensively examined. Early evidence by Lewis (1947) suggested that interpolated activity had to be antagonistic to the criterion response in order to cause forgetting, while Boswell and Bilodeau (1964) and Williams, Beaver, Spence, and Rundell (1969) demonstrated that interpolated activity could cause decrement at recall; however, the majority of research indicated no effect of interpolated activity (Blick & Bilodeau, 1963; Stelmach & Barber, 1970). None of these investigations, however, explicitly manipulated the number of interfering responses as have investigations on prior activity (Ascoli & Schmidt, 1969; Stelmach, 1969; Williams, 1971). Evidence obtained by Melton and Irwin (1940), Thune and Underwood (1943), and Waugh and Norman (1965) in verbal memory studies suggested that the number of interpolated responses could be an important factor in motor STM.

Thus the purpose of this study was to extend previous research in the area and determine the effect of the number of interpolated motor responses on the STM of a simple motor task. It was predicted that interpolated activity would have a retroactive interference (RI) effect; and further, based upon the predictions from verbal memory (Briggs, 1954), it was predicted that the interference effect of interpolated activity would decrease over the retention intervals since the criterion response had longer to spontaneously recover.

METHOD

Subjects.—Thirty-five right-handed male undergraduate students with no prior experience with the apparatus were randomly selected for participation in the study.

Apparatus.—A simple linear motor response was employed. The apparatus consisted of a rectangular slide mounted on three stainless-steel rods 50 cm. long. Attached to the slide was a 5-cm. handle on *S*'s side and a pointer on *E*'s side to indicate the distance moved. The slide could be stopped in either direction when it came in contact with an adjustable stop which could be locked into place. The *E* set the stop by a scale on the front of the apparatus. The distance *S* moved the slide could also be read from this scale.

Experimental procedure and design.—*S*s were blindfolded and wore earphones to exclude extraneous noise. After presentation of a warning buzzer through the earphones, *S*s moved the slide along the track (30 cm.) until it contacted the metal stop defining the criterion position. After contact with the stop for 3 sec., *E* gave the command, "Release."

---

[1] Requests for reprints should be sent to William G. Davenport, Department of Psychology, University of British Columbia, Vancouver 8, B. C.

The *Ss* released the slide starting the retention interval of 60 sec. or 90 sec. The *E* returned the slide to the starting position which remained constant throughout all conditions.

Within each retention interval one of three interpolated learning conditions was utilized: zero, two, or four interpolated learning trials, (0IL), (2IL), or (4IL), respectively. In the 0IL condition, *S* performed three-column addition over the entire retention interval. In the other two groups, the first part of the retention interval involved *S* performing either two or four interpolated trials, while in the remainder of the interval *S* performed three-column addition.

The interpolated trials involved moving to either two (2IL) or four (4IL) of four interpolated positions, two of which were 5 cm. and 10 cm. above the criterion; two were the same distances below the criterion. The order of presentation of the interpolated position was randomized for each *S*.

At recall following the retention interval, *Ss* were required to reproduce both the interpolated positions and the criterion position without the aid of the stop, beginning with the criterion position. This procedure of reproducing all the positions was used to ensure that *S* did not disregard the interpolated traces, removing them from STM, and rendering them incapable of causing interference (Bjork, LaBerge, & Legrand, 1968). Thus *Ss* were concerned with recalling all positions accurately, although only response error in the criterion response was considered in the analysis.

Five *Ss* were randomly assigned to each of 6 treatment conditions in a 2×3 (2 Retention Intervals × 3 Interpolated Positions) factorial design (Winer, 1971). Absolute and algebraic errors in millimeters were the dependent measures.

RESULTS

Absolute error.—A two-way analysis of variance was performed on recall scores. There were significant effects for interpolated trials, $F\ (2,\ 24) = 8.71,\ p < .01$, and retention intervals, $F\ (1,\ 24) = 7.89$, $p < .01$, but no significant interaction between them.

A Duncan multiple-range test was computed on the means of the six treatment conditions. At both the 60- and 90-sec. retention intervals, the 0IL and 2IL groups differed significantly from the 4IL groups. Also in the 4IL group there was significant forgetting over the retention intervals, a result not observed in the 0IL and 2IL groups.

Some decrement in the 0IL group was observed over the retention intervals. Thus, in order to assess when forgetting began, a separate group of five *Ss* was

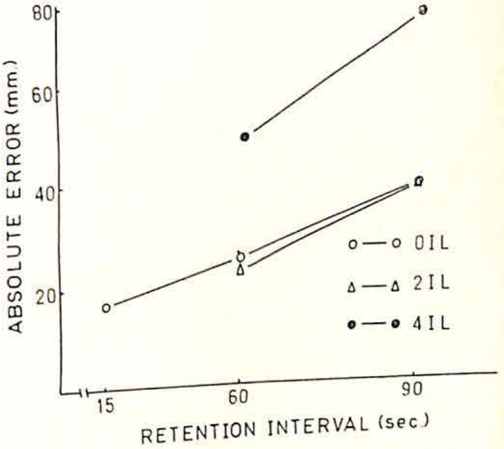


FIG. 1. Recall accuracy as a function of interpolated responses and retention intervals.

tested in a condition of no interpolated trials following a 15-sec. retention interval. A separate analysis was performed on this condition. There was no significant difference in error among the three retention intervals, $F\ (2,\ 12) = 1.21,\ p > .05$.

Algebraic error.—A two-way analysis of variance was performed on the recall scores for algebraic error, but no significant results were obtained. This could indicate that randomizing shorter and longer interpolated movements resulted in a cancellation, to some extent, of algebraic error.

Figure 1 shows the significant interference effect incurred in the 4IL condition.

DISCUSSION

The primary source of forgetting evidenced in this study seems to be due to interference with very little support for trace decay. This concurs with the findings of Williams (1971). There was some evidence of decreased recall accuracy over the retention interval, suggesting that trace decay might be occurring. Consequently, a third 0IL group was tested at a 15-sec. retention interval. Although recall accuracy was better in this condition, a separate analysis of the data on the three 0IL conditions revealed that the decrement in accuracy over the intervals was not significant.

A rival hypothesis which may challenge the interference interpretation concerns the relative attention demands of the three interpolated conditions. Schmidt and Ascoli (1970) indicated that the earlier findings of

Ascoli and Schmidt (1969) and Stelmach (1969) concerning PI effects of prior positions might be interpreted in terms of the differential attentional demands of the two and four prior learning conditions since all groups rested over the retention interval. In this study, however, an attempt was made to approximately equalize the attentional demands of all the conditions by having all groups perform three-column addition, either during the entire retention interval (0IL) or during the remainder of the interval following the interpolated responses.

Pepper and Herman (1970) have shown that mental activity during the retention interval could result in interference which would be reflected in a tendency toward positive algebraic error. This effect of mental activity on algebraic error was not apparent in this study in that in the control condition (0IL) there was a marked tendency toward negative algebraic error. Further, if mental activity caused interference, allowing mental activity for a greater time should probably cause increased interference, and so, greater error. However, there was no significant increase in absolute or algebraic error in the 0IL condition as the length of the retention interval increased. These results for the 0IL condition were assumed consistent for mental activity in the 2IL and 4IL conditions, and thus, the effect of mental activity appeared to be the prevention of rehearsal.

The interference interpretation of forgetting in this study could be confounded by the temporal location of the IL. Patrick (1971) reports that the temporal location of the IL may be more important than the number. He found that approximately the same error resulted from two conditions of IL: one involving five IL and the other involving one IL which occurred at the same temporal location as the fifth IL in the first condition. He postulates this result to be due to a recency effect. Upon examination of both the 2IL and 4IL conditions, it would seem that, to be consistent with the recency hypothesis, there should have been greater forgetting in the 60-sec. than the 90-sec. retention interval due to the closer temporal proximity to recall of the last IL in the 60-sec. retention interval. As this result did not occur, the implication from Patrick's study that temporal location of the last IL in the 4IL condition would cause greater interference than in the 2IL condition due to its closer temporal proximity to recall does not seem warranted.

The data trends indicated that the mean absolute error and variance were very similar for the 0IL and 2IL conditions. From the 0IL and 2IL to the 4IL condition, however, there was a significant increase in absolute error with a corresponding decrease in variance. This result seems to argue against an interpretation such as Stelmach's (1969) that, in the retention of motor tasks, variance and, hence, absolute error increase with conditions which produce interference.

Although interference appears to hold the greater explanatory power, the results obtained only partially support the predictions made concerning the effect of interpolated activity. First, the interference effect of interpolated trials did not constitute a major source of forgetting until the 4IL trials. This supports the results of other investigations in motor STM (Ascoli & Schmidt, 1969; Williams, 1971).

Second, the prediction concerning the decreased interference over the retention interval was not supported. Indeed only in the 4IL condition was the recall accuracy significantly poorer at the 90-sec. retention interval. Thus, it appears that interference increased in the longer retention interval.

These results tended to point out the basic similarity between the verbal and motor STM systems in that interference appeared to constitute the major source of forgetting with little explanatory power afforded to trace decay. However, two possible divergencies between the verbal and motor systems were observed. First, this study confirmed other investigations indicating that interference in motor STM did not occur until at least four interfering responses were present (Ascoli & Schmidt, 1969; Stelmach, 1969; Williams, 1971). Research in verbal STM has indicated that just one interfering item could cause a decrement at recall (e.g., Keppel & Underwood, 1962).

Second, RI usually decreases over retention intervals in verbal memory, but in this study increased over the retention interval. Possibly this increase was due to the fact that the criterion response had not spontaneously recovered to any significant degree over the retention interval. Indeed, this may be the case since in verbal learning, the extinction hypothesis concerning spontaneous recovery of the criterion response was tested primarily in secondary memory in which the retention intervals involved periods of up to several days.

In conclusion, these results indicated that interference caused by interpolated activity was a major source of forgetting in motor STM. Further research should be carried out to establish the generality of this effect over other types of motor responses.

REFERENCES

Ascoli, K. M., & Schmidt, R. A. Proactive interference in short-term motor retention. *Journal of Motor Behavior*, 1969, **1**, 29–36.

Bjork, R. A., LaBerge, D., & Legrand, R. The modification of short-term memory through instructions to forget. *Psychonomic Science*, 1968, **10**, 55–56.

Blick, K. A., & Bilodeau, E. A. Interpolated activity and the learning of a simple skill. *Journal of Experimental Psychology*, 1963, **65**, 515–519.

Boswell, J. J., & Bilodeau, E. A. Short-term retention of a simple motor task as a function of interpolated activity. *Perceptual and Motor Skills*, 1964, **18**, 227–230.

Briggs, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *Journal of Experimental Psychology*, 1954, **47**, 285–293.

Keppel, G., & Underwood, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

Lewis, D. Positive and negative transfer in motor learning. *American Psychologist*, 1947, **2**, 423. (Abstract)

Patrick, J. The effect of interpolated motor activities in short-term motor memory. *Journal of Motor Behavior*, 1971, **3**, 39–48.

Pepper, R. L., & Herman, L. M. Decay and interference effects in the short-term retention of a discrete motor act. *Journal of Experimental Psychology*, 1970, **83**(2, Pt. 2).

Melton, A. W., & Irwin, J. M. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *American Journal of Psychology*, 1940, **53**, 173–203.

Schmidt, R. A., & Ascoli, K. M. Attention demand during storage of traces in motor short-term memory. *Acta Psychologica*, 1970, **34**, 497–504.

Stelmach, G. E. Prior positioning responses as a factor in short-term retention of a simple motor task. *Journal of Experimental Psychology*, 1969, **81**, 523–526.

Stelmach, G. E., & Barber, J. L. Interpolated activity in short-term motor memory. *Perceptual and Motor Skills*, 1970, **30**, 231–234.

Thune, L. E., & Underwood, B. J. Retroactive inhibition as a function of degree of interpolated learning. *Journal of Experimental Psychology*, 1943, **32**, 185–200.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

Williams, H. L., Beaver, W. S., Spence, N. T., & Rundell, O. K. Digital and kinesthetic memory with interpolated information processing. *Journal of Experimental Psychology*, 1969, **80**, 537–544.

Williams, I. D. The effects of practice trials and prior learning on motor memory. *Journal of Motor Behavior*, 1971, **3**, 205–211.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw–Hill, 1971.

(Received February 23, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 138–141

ON THE INHIBITORY EFFECT OF A SECOND STIMULUS FOLLOWING THE PRIMARY STIMULUS TO REACT:

A SUCCESSFUL REPLICATION

FREDERICK L. KITTERLE[1] AND HARRY HELSON[2]

University of Massachusetts

H. Helson and J. A. Steger found in 1962 that a second stimulus (S_2) interfered with the reaction time to a primary stimulus (S_1). Reaction time (RT) was found to be a curvilinear function of the stimulus-onset-asynchrony (SOA) between 80 and 100 msec. The failure of other investigators to obtain this **U**-shaped function was interpreted as being due to differences in procedure and methodology, the chief difference being the spatial separation between S_1 and S_2.

The purpose of this study was to resolve discrepancies between the results reported by Helson and Steger (1962) and subsequent attempts to replicate the study by Koplin, Fox, and Dozier (1966), Lappin and Eriksen (1964), and Herman and Kantowitz (1969). Briefly, Helson and Steger found that a second stimulus (S_2) which followed a primary stimulus (S_1) lengthened the reaction time (RT) to S_1, as shown by an inverted **U**-shaped curve, which described RT as a function of stimulus-onset-asynchrony (SOA). Maximal interference in RT occurred when S_2 onset was 80 to 100 msec. later than the S_1 onset.

In a number of attempted replications of this study the inverted **U**-shaped function was not found (Herman, 1969; Herman & Kantowitz, 1969; Herman & McCauley, 1969; Koplin et al., 1966; Lappin & Eriksen, 1964). These attempted replications have differed from the original Helson and Steger study in a number of ways that might have significantly influenced the results of the experiments.

One of the major differences of these experiments was in the display used. Helson and Steger (1962) used a three-light stimulus display. In their experiment, both end lights functioned as S_1 and the middle light functioned as S_2. Over trials, the position of S_1 with respect to S_2 varied, i.e., from the left to the right. However, in the Lappin and Eriksen (1964) study, a two-light display was used. The lights labeled S_1 and S_2 were randomly varied over trials; S_1 was responded to on 50% of the trials, while in the Helson and Steger experiment no overt response was required to S_2.

Herman (1969) investigated the effects of stimulus display on RT to S_1 in a more extensive experiment. One group (Group 1) responded to one of two stimulus lights as in the Lappin and Eriksen (1964) study. Another group (Group 2) viewed three lights arranged in a triangle with S_2 at the apex. The bottom two lights forming the base of the triangle were the S_1 lights. The conditions of this group were similar to those of Helson and Steger (1962). A third group responded to S_1 alone. The results of this experiment showed that RT was independent of interstimulus interval (ISI) for all groups. Close examination of their stimulus display reveals a spatial distance of 5.65 cm. between S_1 and S_2, while the distance in the Helson and Steger study was only 1.90 cm. so the bowing effect may be dependent upon some critical distance between S_1 and S_2, a supposition we have tested in this study.

The purpose of this investigation is to determine whether the inverted **U**-shaped RT function depends upon S_2 being neutral (i.e., no response is made to it) and whether there is a critical distance between S_1 and S_2 for the Helson–Steger effect to appear.

---

[1] Requests for reprints should be sent to Frederick L. Kitterle, Department of Psychology, University of Toledo, Toledo, Ohio 43606.

[2] Now at 15 The Crescent, Berkeley, California 94708.

Experiment I

Method.—Ten *Ss* in the introductory psychology course at the University of Massachusetts served as *Ss* in the experiment in order to fulfill part of the requirements for course credit.

The apparatus consisted of a vertical panel board, 45.7 × 61.0 cm., on which were mounted horizontally three neon lamps (GE NE-15) behind ground-glass apertures 1.75 cm. in diameter. The lamps were mounted in the center of the panel and spaced 1.90 cm. apart. The lamps were mounted in a light housing and shielded so that each stimulus was lighted independently of the others. Above the center lamp (S_2) was a red cross serving as a fixation point. The response button, a Western Union telegraph key, was mounted in the middle of a slightly angular horizontal panel 22.9 cm. from the vertical panel. Both the horizontal and the vertical panels were painted in a flat black.

The *E* sat behind the panel board and by pressing a switch could turn on the stimulus light to the left or right of S_2. A Hunter interval timer controlled the ISI between the onset of S_1 and S_2. The ISIs ranged from 20 to 200 msec. in steps of 20 msec. The 0-msec. ISI was the control condition in which no S_2 was presented. Both S_1 and S_2 stayed on until *S* responded. Response time in milliseconds was recorded by a Hunter Klockcounter. The experiment was run in a normally lighted room. The lights which served as stimuli were distinctly visible against the black background.

As in an earlier study (Helson & Steger, 1962), the visual stimuli were counterbalanced; i.e., the primary stimulus (S_1) was either on the right or the left of the middle or interpolated stimulus which always served as S_2. The procedure is outlined in the instructions given to the *Ss*.

TABLE 1

Theoretical and Observed Reaction Times to a Primary Stimulus (S_1) followed at Various Primary Stimulus-Onset-Asynchronies (SOAs) Stimulus-Onset-Asynchronies (SOAs) by a Second Stimulus (S_2).

SOA	Theoretical value	Observed value
No S_2	171.86	172.75
20	175.06	173.09
40	177.46	177.35
60	179.06	178.59
80	179.86	182.05
100	179.86	181.64
120	179.06	176.07
140	177.46	177.51
160	175.06	174.77
180	171.86	172.57
200	167.86	168.85

Note.—The theoretical values were based on the following equation obtained by least-squares criterion: $Y = 171.86 + .18X - .001X^2$. Reaction times are measured in milliseconds.

TABLE 2

Analysis of Variance of the Results of Experiment I

Source of variation	*df*	*MS*	*F*
SOA	10	162.60	17.81*
lin SOA	1	127.57	*ns*
quad SOA	1	1271.08	13.92*
residual	8	28.41	*ns*
Ss (S)	9	4950.60	
S × SOA	90	91.3	

Note.—The scores used were the means of the reaction time per *S* per stimulus-onset-asynchrony (SOA). Data analysis was based on 1,100 observations (10 *Ss* × 11 SOAs × 10 replications).

* $p < .001$.

> This is an experiment in simple reaction time. You are to respond to the light *as quickly as you possibly can* by lifting your finger (index) from the response key. Note the lights facing you. I will tell you which end light you are to watch. If I say "left" you are to watch the light on the left and react to it *as quickly as possible.* The procedure will be this: I will say "right" or "left" and you are to then depress the response key. Then I will say "ready." A short time later the light you are to watch will come on. As soon as it comes on you are to *react to it as quickly as possible by lifting your finger from the response key.*

The left and right presentations of S_2 were randomly distributed and counterbalanced so that at each of the 11 ISIs there were five right and five left presentations of S_1 with respect to S_2. The primary stimulus, S_1, appeared after variable intervals ranging from .5 to 3 sec. after the ready signal. Each *S* served as his own control in that his average RT without S_2 served as his zero or base line. The ISIs were presented in random order. The *Ss* were given four practice trials in which they responded to S_1 alone in order to familiarize them with the procedure. The *Ss* were not told the purpose of the experiment or the role of S_2 until all of the data of the experiment were gathered.

The data for this study were based on 1,100 observations (10 *Ss* × 11 ISIs × 10 Replications).

Results.—The results of Exp. I are presented in Table 1. The data indicated that RT to S_1 is not flat over ISI but follows a parabolic curve which peaks at 80–100 msec., as in the Helson–Steger (1962) study. This conclusion is verified by the significant quadratic effect, F (1, 90) = 13.92, $p < .001$, in the analysis of variance presented in Table 2.

The results of this analysis support the conclusions of Helson and Steger (1962) that the RT to a primary stimulus is

TABLE 3

RESULTS OF EXPERIMENT II

SOA	Cond. 1	Cond. 2	Cond. 3
No S_2	210.57	181.00	207.00
20	210.93	175.00	196.80
40	207.40	191.83	200.87
60	226.77	190.30	198.00
80	250.03	192.23	195.37
100	235.37	210.42	190.67
120	223.93	193.27	196.77
140	214.20	179.93	205.13
160	222.13	182.37	194.33
180	201.27	187.17	205.20
200	207.27	193.17	198.73

Note.—Each entry is the mean of three *Ss'* reaction time per condition per SOA. Condition 1 is the regular Helson-Steger three-light display; Cond. 2, a two-light display with 1.90 cm. separation between S_1 and S_2; Cond. 3, a two-light display with a separation of 3.81 cm. between S_1 and S_2.

lengthened when it is followed by a secondary stimulus under their conditions.

EXPERIMENT II

The purpose of this experiment was to investigate the effects of S_2 on S_1 in a two-light stimulus display when competition was held to a minimum by blocking trials, so that the spatial position of S_1 was held constant for a given number of trials.

Method.—Three *Ss* from the introductory psychology course participated in this experiment as part of the course credit.

The apparatus of this experiment was similar to that used in Exp. I. The right-most light of the display was covered with black tape leaving a two-light stimulus display with an interlight separation of 1.90 cm. In another condition (Cond. 3), the middle light of the three-light display was blacked out leaving a two-light display with a separation of 3.81 cm.

TABLE 4

ANALYSIS OF VARIANCE OF THE RESULTS OF EXPERIMENT II

Source of variation	*df*	*MS*	*F*
Cond. (C)	2	8607.07	*ns*
SOA	10	212.51	*ns*
Ss (S)	2	1706.61	
C × SOA	20	391.97	2.16*
C × Linear SOA	1	225.96	*ns*
C × Quadratic SOA	1	947.31	5.22*
C × S	4	10904.06	
S × SOA	20	206.12	
C × S × SOA	40	168.91	

Note.—This analysis is based on the mean of three *Ss'* judgments per condition per stimulus onset asynchrony (SOA).
* $p < .01$.

Experimental conditions were presented to *Ss* in a counterbalanced order. Condition 1 was a straight replication of Exp. I. In Cond. 2, *Ss* responded to the left light of a two-light stimulus display for 10 trials and then the right light for another 10 trials. Cond. 3 was similar to Cond. 2 except that S_1 and S_2 were separated by 3.81 cm.

The primary stimulus (S_1) appeared after variable intervals ranging from .5 to 3 sec. after the ready signal. Each *S* served as his own control in that his average RT without S_2 served as his zero or base line. The *Ss* were given four practice trials in which they responded to S_1 alone in order to familiarize them with the procedure. The *Ss* were not told the purpose of the experiment or the role of S_2 until all of the data of the experiment were gathered.

Results.—The results of this experiment are presented in Table 3, and an analysis of variance of these results is presented in Table 4. The significant first-order interaction between ISI and conditions indicates that each condition differentially affected the forms of the RT function over ISIs. Further analysis reveals a significant quadratic ISI × Conditions interaction, $F(1, 60) = 5.22$, $p < .001$. In general, Cond. 1 and Cond. 2 produced parabolic curves, while the RT was virtually flat over ISI for Cond. 3.

DISCUSSION

The results of these experiments indicate that the **U**-shaped function reported by Helson and Steger (1962) is dependent upon at least two stimulus parameters. The first of these is the neutrality of S_2. If, as in the Lappin and Eriksen (1964) study, S_2 is responded to on some trials, the interfering effect of S_2 on S_1 is diminished (Herman & Kantowitz, 1969). However, if the trials are blocked so that S_2 is not responded to or if a three-light display is used, the inverted **U**-shaped RT function is obtained.

A second important parameter is the distance between S_1 and S_2. A change in spatial separation from 1.90 cm. to 3.81 cm. nullified any effect of S_2 on S_1.

Thus, it is evident that the failures to replicate Helson and Steger (1962) may be due to failure in minimizing response competition and in keeping the S_1–S_2 separation within a critical distance. Further studies are needed to specify the nature and locus of the S_2 interference.

REFERENCES

HELSON, H., & STEGER, J. A. On the inhibitory effect of a second stimulus following the primary stimulus to react. *Journal of Experimental Psychology*, 1962, **64**, 201–205.

HERMAN, L. M. Effects of second signals on response time to first signals under certainty and uncertainty. *Journal of Experimental Psychology*, 1969, **80**, 106–112.

HERMAN, L. M., & KANTOWITZ, B. H. Effects of signals occurring after response selection on responses to first signals. *Journal of Experimental Psychology*, 1969, **80**, 570–572.

HERMAN, L. M., & MCCAULEY, M. E. Delay in responding to the first stimulus in the "psychological refractory period" experiment: Comparisons with delay produced by a second stimulus not requiring a response. *Journal of Experimental Psychology*, 1969, **81**, 344–350.

KOPLIN, J. H., FOX, R., & DOZIER, R. A failure to replicate the inhibitory effects of a second stimulus following the primary stimulus to react. *Journal of Experimental Psychology*, 1966, **72**, 914–916.

LAPPIN, J. S., & ERIKSEN, C. W. Inhibition of a simple visual reaction time by a second stimulus: A failure to replicate. *Psychonomic Science*, 1964, **1**, 293–294.

(Received February 24, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 142–146

IRRELEVANT INFORMATION, IRREGULARITY, AND THE PACING OF REHEARSAL IN SEQUENTIAL SHORT-TERM MEMORY[1]

ROBERT KARSH[2] AND RICHARD A. MONTY

Human Engineering Laboratory, Aberdeen Proving Ground, Maryland

The *S*s were required to keep track mentally of the number of occurrences of each of four different symbols presented sequentially. It was found that when irrelevant information which closely resembled relevant information was introduced into the sequence performance was disrupted. Delaying identification of relevancy further disrupts performance. The disruption was not attributable to the resulting irregularity of presentation. It was further found that the detrimental effects of irregular presentation rates can be eliminated by insertion of cues enabling *S*s to pace rehearsals of their mental tallies.

When *S*s are required to tally mentally the number of occurrences of each of several different symbols or letters presented repetitively in a sequence, the majority of *S*s mentally picture a set of windows, one corresponding to each category of symbols presented. Monty (1968) and Monty, Taub, and Laughery (1965) have demonstrated that individual running tallies are then kept in each window and only the tallies are rehearsed after the presentation of each successive symbol. Monty, Karsh, and Taub (1967a) have found this task to be remarkably resistant to interference from irrelevant information inserted into the sequence. Specifically, when "irrelevant letters" (U, V, and W—to be ignored by *S*) were presented at random as part of a periodic sequence of both irrelevant and "relevant letters" (Q, R, and S—to be tallied by *S*), the irrelevant letters failed to disrupt keeping-track performance. Further, they actually produced better performance than that of a control group in which the irrelevant letters were omitted from the sequence, thereby resulting in an aperiodic or irregular presentation of the relevant letters only. On the average, performance was better than that of a neutral control group which viewed letters at a regular rate of presentation. Subsequently, Monty, Karsh, and Taub (1967b) demonstrated that a brief flash of light introduced just prior to each successive symbol in a periodic sequence did enhance performance relative to a control group with no flash of light apparently by serving as a cue which enables *S* to complete ongoing rehearsal of the current state of the information in preparation for the next stimulus in the sequence. Presumably, the irrelevant information used in the earlier study (Monty et al., 1967a) served the same purpose.

In short then, it appears that when given irrelevant and relevant information intermixed in a sequence, *S*s can easily make a perceptual sort in a manner that does not interfere with either the encoding or rehearsal of the relevant information. Would the same result occur if the irrelevant information closely resembled the relevant information? Would the presence of irrelevant information within a sequence disrupt the rehearsal process if *S* were not told whether a letter was relevant or not until some time after it was shown, thus forcing him to attend to it and perhaps enter it into memory (but not necessarily to add it to his running tally)? Can the disruptive effects of irregular presentation be affected by the insertion of a cue with which *S* can pace rehearsal? These were the major questions to which the present experiment was addressed.

---

[1] The authors gratefully acknowledge the assistance of Jean Standiford in reducing the data.

[2] Requests for reprints should be sent to Robert Karsh, Behavioral Research Directorate, Human Engineering Laboratory, Aberdeen Proving Ground, Maryland 21005.

METHOD

Subjects.—One hundred and fifty-four students from Harford Community College, Churchville, Maryland, were paid volunteers in this experiment.

Apparatus and procedure.—The apparatus has been described more fully elsewhere (Monty et al., 1965). Briefly, *S* was seated in front of an Industrial Electronic Engineers, Inc., rear-projection display. Sequences of letters seen as an outline of light against a dark background were preprogrammed with a Grason-Stadler SCAT system and projected one letter at a time onto the 3.8 × 4.8 cm. frosted-glass display surface. The *S*'s task was to observe and count mentally the number of occurrences of each of four relevant letters (Q, R, S, and T) and to record their tally in writing at the end of each sequence. A sequence is defined as the *total* number of relevant and irrelevant letters presented on a trial.

The *S*s were randomly assigned to seven groups of 22 each as illustrated in Fig. 1. The four irrelevant information groups were designed to determine (*a*) whether irrelevant information which looked identical (also Q, R, S, and T) to relevant information would disrupt performance and (*b*) if forcing *S* to enter the material into memory before providing information pertaining to relevancy would seriously disrupt keeping-track performance. This was accomplished by earmarking each letter in a sequence with either a red or green light of .3-sec. duration. The light was projected in the same manner as the letters and appeared as a 2.54-cm. square. The *S*s were instructed to keep track of only those letters associated with a green light and to ignore all letters associated with a red light. The four groups were differentiated on the basis of where, during the 1-sec. off-time between letters, the red or green light appeared. In Group IP (irrelevant, preceding), the color occupied the last .3 sec. of each off-time and immediately *preceded* its associated letter. In group IF1 (irrelevant, following), the color occupied the first .3 sec. of each off-time and immediately *followed* the letter with which it was associated. In Group IF2, the color occupied the middle .3 sec. of each off-time and followed its letter, and in Group IF3, it occupied the last .3 sec. of each off-time and followed its associated letter. These last three groups, then, force *S*s to discard information during the period normally used for rehearsal.

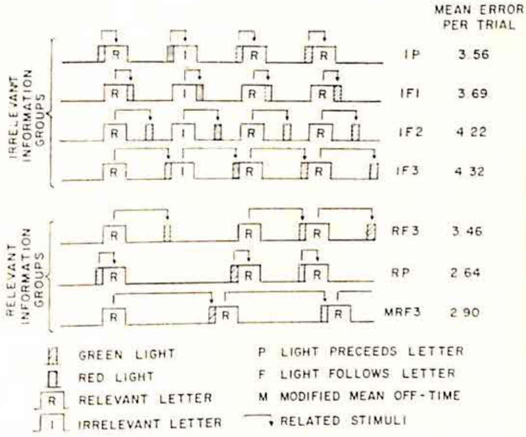


FIG. 1. Representation of the seven test groups and their mean errors per trial.

The total number of relevant letters within any given sequence was defined as trial length. All *S*s were presented with trial lengths of 8, 12, 16, and 20 relevant letters. So that *S*s would not learn how much of the information was irrelevant, three levels of irrelevancy were included. In any given sequence either none of the letters were irrelevant, $\frac{1}{3}$ of the letters were irrelevant, or $\frac{1}{2}$ of the letters were irrelevant. The levels of irrelevancy will be referred to as conditions. In Cond. 1, no irrelevant letters were shown. The presentation rate was thus 1 relevant letter every 1.5 sec. In Cond. 2, the number of irrelevant letters shown was equal to $\frac{1}{2}$ of the number of relevant letters per sequence. This resulted in a mean presentation rate of 1 relevant letter every 2.25 sec. In Cond. 3, the number of irrelevant letters shown was equal to the number of relevant letters per sequence. This resulted in a mean presentation rate of 1 relevant letter every 3.0 sec.

Each trial length was paired with each condition for a total of 12 combinations. The 12 combinations for each group were presented in random order during each of four blocks of 12 sequences. The orders in which the letters within a sequence were presented were chosen at random with the restriction that each relevant letter and each irrelevant letter (when called for) appeared at least once in every sequence. The *S*s were instructed that each letter would always appear but had no knowledge of the total number of relevant or irrelevant letters to be presented in each sequence.

The resulting effect was that *S*s were presented with a sequence of from 8 to 40 periodically occurring letters, some relevant and some irrelevant as indicated by the red or green light associated with each letter. The rate was always 1 letter every 1.5 sec., with a .5-sec. on-time and a 1-sec. off-time. The relevant and irrelevant letters within a sequence were randomly distributed. No restrictions were imposed on the number of irrelevant letters that might occur between relevant letters, other than those imposed by the level of irrelevancy and the trial length. Thus, the relevant letters were shown aperiodically in all sequences except those composed of Cond. 1.

Three additional groups of *S*s were given relevant information only and served primarily as controls for the irrelevant information groups. Two of these groups were designed to investigate the effects of aperiodic relevant letter presentations both within a sequence and from sequence to sequence, in the absence of irrelevant information. In Groups RP (relevant, preceding) and RF3 (relevant, following) the identical sequences were presented as in the

Irrelevant Information Groups IP and IF3, respectively, with the exception that the irrelevant letters and their associated red lights were switched off and thus, never appeared. The resulting effect was that the remaining relevant letters appeared at irregular intervals of time for those sequences which had originally contained irrelevant letters (Cond. 2 and 3).

The Relevant Information Only Group MRF3 (modified, relevant, following) served as a control for the irregular presentation patterns of the aforementioned two groups. Here again, the relevant letter sequences were identical to those used in the irrelevant information groups and again, no irrelevant letters or red lights appeared. However, the intervals at which the relevant letters appeared were modified so that the mean presentation rates of 1.5 sec., 2.25 sec., and 3.0 sec. employed in the above groups became the *actual* presentation rate. In other words, all letters appeared at regular intervals within any given sequence but at a different rate from sequence to sequence.

In all cases, *S*s were informed of the temporal relationship between the letters and the lights, (i.e., whether the light followed or preceded the letter). No suggestion was made, however, that the light might be useful to them except to differentiate between relevant and irrelevant letters.

All seven groups were given 2 practice sequences to ensure that *S*s understood the task. These were followed by 48 experimental sequences with a 15-sec. intersequence interval for recording answers. A 1,200-Hz. tone signaled the beginning and end of each sequence.

RESULTS

Absolute error, defined as the sum of the absolute differences without regard to sign, between the number of each relevant letter presented in a sequence and the number reported, served as the measure of performance. For example, if four Qs, three Rs, one S, and four Ts were presented and *S* reported five Qs, one R, one S, and three Ts, the number of errors for that sequence would have been one for the letter Q, two for the letter R, and one for letter T, for a total of four. Because of differential learning effects early in the experiment only the data for the fourth block are reported here. The absolute error scores for each Group × Trial Length combination were subjected to an analysis of variance (Butler, Kamlet, & Monty, 1969) with groups as a between-effect and trial length as a within-effect.

The main effect for trial length was significant, F (3, 441) = 134.36, $p < .01$, indicating that the longer the trial length the poorer the performance (mean errors, respectively, of 2.13, 3.18, 3.93, and 4.94 for trial lengths of 8, 16, and 20). Of greater interest is the significant main effect for groups, F (6, 147) = 3.00, $p < .01$. The means underlying the effect are shown in Fig. 1. Finally, the Groups × Trial Length interaction failed to reach significance at the .05 level of significance.

DISCUSSION

Does the inclusion of irrelevant information which is similar to relevant information disrupt keeping-track performance? It can be seen from Fig. 1 that Group IP made an average of 3.56 errors per trial in contrast with its control for irregularity of relevant letter presentation, Group RP, with only 2.64 errors per trial. Thus, unlike the earlier experiment in which the relevant and irrelevant materials were dissimilar (Monty et al., 1967a), irrelevancy in the present study led to disruption of keeping-track performance. Similarity of materials, then, appears to be a very important factor. This finding lends support to the contention of Erlick (1962) and Hodge (1959) to the affect that the ability to filter out irrelevant material (noise) in complex tasks increases as a function of the ease of differentiating or categorizing the relevant from the irrelevant aspects of the task and that an important requirement for irrelevancy to have any detrimental effect is that the irrelevant information be at other times relevant.

Does a delay in identifying the relevancy of information further disrupt performance? Underwood's (1952) reasoning regarding the need for contiguity of stimulus and response would lead one to predict yes since the greater the delay of information, the more *S* must rely on memory to maintain the information gap. It can be seen from Fig. 1 that when stimuli were identified immediately following their presentation (IF1), performance was nearly identical to when identification immediately preceded (IP) each stimulus (mean errors per trial, respectively, of 3.69 vs. 3.56). However, as the identification moved further away from the stimulus, forcing *S* to enter the stimulus in memory and delay rehearsal, performance deteriorated gradually (mean errors per trial of 4.22 and 4.32 for Groups IF2 and IF3). In short, the longer the delay the poorer the performance. It cannot be stated unequivocally, however, that the disruption is due to the delay per se. It must be noted that as the delay increases, the location of the cue gets closer to the next stimulus so that *S* is also required to make a decision about the old cue at the same time he is processing the new letter. Thus, the disruption may reflect the difficulty of assessing the previous stimulus and processing the new one at the same time. Cahill and Hovland (1960) have termed errors of this latter type perceptual inference (PI) errors as contrasted with errors of the first type, memory (M) errors. This confound noted between the two types of errors in the present experiment is logically similar to that encountered by Bourne, Goldstein, and Link (1964) in their work on concept learning. Both Bourne et al. and Cahill and Hovland, however, are led to conclude that only a small proportion of errors in such situations are attributable to improper perceptual inferences. More errors result from *S*s' apparent failure to remember earlier stimuli.

There is, however, another technique *S*s could have used to minimize disruption of performance, that is, to rehearse only the remaining three letters *not* presented while awaiting identification of the last letter. For example, if an R were presented, *S* could rehearse the previous running tallies of Qs, Ss, and Ts while awaiting identification of the R. By contrast, because of the evidence presented by Monty, Wiggins, and Karsh (1969), demonstrating that subtractions are substantially more difficult to perform than additions, it is highly unlikely that *S*s tallied each letter at the time of presentation and then subtracted from the tally if the letter proved irrelevant. If this were the technique employed, a much greater level of error would have been predicted. Further, the disruption cannot be accounted for simply on the basis of an irregular rate of presentation as indicated by the fact that the Control Group RF3 generated only 3.46 errors per trial as compared with 4.32 for Group IF3.

What about the role of irregularity per se? It will be recalled that an earlier investigation by Monty et al. (1967b) demonstrated that with a regular rate of presentation, a brief flash of light (similar to that employed here) introduced just prior to each successive symbol enhanced performance presumably by cuing *S*s to terminate the multiple rehearsals postulated by Monty et al. (1967a), in preparation for the next stimulus. There is strong evidence here that the light can also be used to offset the disruptive effects of irregular rates of presentation noted by Monty et al. (1967a). Notice, for example, that Group MFR3 in which presentation rate was regular and Group RP in which it was irregular committed virtually the same number of errors per trial (mean errors per trial, respectively, of 2.90 vs. 2.64). This similarity is attributed to the fact that each letter was preceded by a brief flash of light (except in the case of the first letter for Group MFR3) which presumably serves to cue *S*s to complete rehearsal of the current state of the information in preparation for the next stimulus in the sequence (Monty et al., 1967b). By contrast, when the cue did not always immediately precede the letter (RF3), performance deteriorated slightly to 3.46 errors per trial. Apparently, *S*s are either stopping rehearsal at the time of the cue and forgetting their tallies during the interval before the next letter or attempting to rehearse during that interval and getting interrupted by the next letter. In short, a cue which immediately precedes each successive letter is better than one which follows each successive letter by some fixed interval.

In summary, it appears that irrelevant information which bears a close similarity to relevant information can disrupt keeping-track performance, that a delay in identifying the relevancy further disrupts performance either through PI errors, M errors, or both, and that irregular rates of presentation do not disrupt performance provided that *S* is cued to the pending onset of the next symbol.

REFERENCES

Bourne, L. E., Jr., Goldstein, S., & Link, W. E. Concept learning as a function of availability of previously presented information. *Journal of Experimental Psychology*, 1964, **67**, 439–448.

Butler, D. H., Kamlet A. S., & Monty, R. A. A multi-purpose analysis of variance FORTRAN IV computer program. *Psychonomic Monograph Supplements*, 1969, **2**, 301–319.

Cahill, H. E., & Hovland, C. I. The role of memory in the acquisition of concepts. *Journal of Experimental Psychology*, 1960, **59**, 137–144.

Erlick, D. E. The ability to filter noise from a visual task when the noise and signal are presented sequentially. *Journal of Experimental Psychology*, 1962, **63**, 111–114.

Hodge, M. H. The influence of irrelevant information upon complex visual discrimination. *Journal of Experimental Psychology*, 1959, **57**, 1–5.

Monty, R. A. Spatial encoding strategies in sequential short-term memory. *Journal of Experimental Psychology*, 1968, **77**, 506–508.

Monty, R. A., Karsh, R., & Taub, H. A. Keeping-track of sequential events: Irrelevant information and paced rehearsal. *Perceptual and Motor Skills*, 1967, **24**, 99–103. (a)

Monty, R. A., Karsh, R., & Taub, H. A. Pacing of rehearsal in sequential short-term memory. *Journal of Experimental Psychology*, 1967, **74**, 300–302. (b)

Monty, R. A., Taub, H. A., & Laughery, K. R. Keeping-track of sequential events: Effects of rate, categories, and trial length. *Journal of Experimental Psychology*, 1965, **69**, 224–229.

Monty, R. A., Wiggins, H. F., & Karsh, R. Keeping-track of sequential events: Manipulation of the incrementing process. *Journal of Experimental Psychology*, 1969, **80**, 408–411.

Underwood, B. J. An orientation for research on thinking. *Psychological Review*, 1952, **59**, 209–220.

(Received March 4, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 147–151

SHORT-TERM RETENTION OF AUDITORY SEQUENCES AS A FUNCTION OF STIMULUS DURATION, INTERSTIMULUS INTERVAL, AND ENCODING TECHNIQUE[1]

JOHN G. MISCIK, JERALD M. SMITH, NORMAN H. HAMM, KENNETH A. DEFFENBACHER, AND EVAN L. BROWN[2]

University of Nebraska at Omaha

Two experiments tested whether short-term recall for auditory digit sequences could be improved by efficient encoding techniques and increases in either digit duration or interdigit interval. All three hypotheses received strong support from the data. In addition, analysis of interactions between length of retention interval and digit duration, interdigit interval, and encoding technique led to the conclusions that longer digit durations and interdigit intervals permit increased resistance to forgetting at longer retention intervals while efficient encoding techniques improve recall regardless of the length of the retention interval.

A number of short-term memory studies have shown that accuracy of recall varies indirectly with stimulus presentation rate (e.g., Aaronson, 1968). Clearly stimulus presentation rate for sequential stimuli depends on both the duration of individual items in the sequence and the interitem interval. In general, short-term memory studies which vary stimulus presentation rate have done so only by manipulating interstimulus interval. Herrington (reported by Bergstrom, 1907, pp. 222–228) did manipulate stimulus duration as well in a task involving serial learning of visually presented nonsense syllables. Herrington found no effect of duration on recall but did find greater recall with increased interstimulus intervals. More recently, Sitterly (1968) varied both stimulus duration and interstimulus interval for visually presented digits and found improved recall for increases of both variables.

Greater retention at longer stimulus durations and interstimulus intervals might be explained as resulting from *S*'s use of the extra time for encoding (organization of material in short-term store) or rehearsal. Generally, efficient encoding techniques such as chunking (Miller, 1956) have been found to facilitate recall. Morganstein (1970), for example, found recall for sequences for colored lights increased with decreases in the length of code used to represent a stimulus pattern.

The present experiments attempt to determine whether Sitterly's (1968) findings for visual digit sequences also hold for auditory ones. Experiments I and II tested whether longer digit durations and longer interdigit intervals would result in greater recall. Experiment II also tested whether efficient encoding techniques would result in superior recall across levels of digit duration and interdigit interval.

EXPERIMENT I

Method

Subjects and design.—Thirty-two University of Nebraska at Omaha students seeking extra credit in an introductory course were randomly assigned to the between-*S*s cells of a 2 (digit durations) × 2 (interdigit intervals) × 4 (sequences of four-digit numbers) × 3 (trial blocks) × 5 (retention intervals) factorial design with repeated measures on the last two factors. Levels of digit duration were .5 sec. and 1 sec., while levels of interdigit interval were 0 sec. and 1 sec. Four different sequences of 60 four-digit numbers were used to increase the

---

[1] Financial support of the University of Nebraska at Omaha Senate Research Committee through a grant to the fifth author is gratefully acknowledged. Experiment II, conducted under the direction of the fourth author, is based on a thesis submitted by the first author in partial fulfillment of the requirements for the master's degree at the University of Nebraska at Omaha. Experiment I, conducted under the direction of the third author, is based on a paper submitted by the second author to the 1971 Nebraska Psychological Association's undergraduate research competition.

[2] Requests for reprints should be sent to Evan L. Brown, Department of Psychology, University of Nebraska, Omaha, Nebraska 68101.

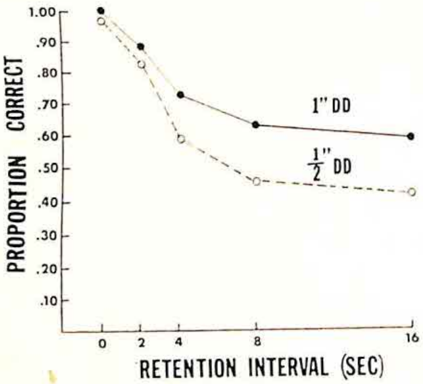


FIG. 1. Digit Duration (DD) × Retention Interval interaction.

generality of the results. Within each of the three blocks of 20 trials *S*s were tested for retention four times at each of five retention intervals, 0, 2, 4, 8, and 16 sec.

Apparatus and materials.—Each four-digit number was selected from a random-number table with the restrictions that none of the digits in a given number occurred twice and zeroes were excluded. Recording of the digit sequences was done at a tape speed of 9.52 cm/sec using a standard two-track monaural tape recorder equipped with a remote start–stop switch. The tapes for the two .5-sec. digit duration conditions involved an additional step in their processing. The .5-sec. digit duration/0-sec. interdigit interval tape was produced by time compressing a 1-sec. digit duration/0-sec. interdigit interval tape by a factor of two; a 1-sec. digit duration/2-sec. interdigit interval tape was similarly time compressed to yield the .5-sec digit duration/1-sec. interval tape. The time compression was accomplished with a speech-time compressor manufactured by Discerned Sound (Model SL14). This speech-time compressor is an electromechanical compressor of the Fairbanks type (Fairbanks, Everitte, & Jaeger, 1954). Time compression was accomplished by discarding alternate time samples which in this study were 38 msec. in length. Speech compressed in this fashion is unaffected with respect to vocal pitch.

A chart containing 12 rows of 10 randomly selected capital letters, excluding the letters I and O, was taped to the wall at eye level approximately 1 m. in front of *S*. Each letter was 2.54 cm. high. These letters were read aloud by *S* during the retention interval to minimize rehearsal, after the manner of Peterson and Peterson (1959). The retention intervals were timed with a standard .1-sec stopwatch.

Procedure.—Each of the 60 trials consisted of the auditory presentation of a four-digit number at the appropriate rate, followed immediately by *S* reading aloud, as rapidly as possible, successive rows of letters until the end of the particular retention interval. The *E* then recorded *S*'s verbal response, and the procedure was repeated. Within every block of 5 trials, *S* was tested once at each retention interval. The order in which retention intervals were tested was randomly determined within each 5-trial block and was constant for all *S*s. Each *S* received 6 practice trials.

Results

An error was scored each time *S* failed to recall a particular digit in its correct serial position. As each retention interval was tested four times within each 20-trial block, the maximum number of errors for these four tests would be 16 (four-digit number × four tests). Each score entered into the previously described analysis of variance design was the proportion correct of 16.

The *S*s averaged .65 correct in the .5-sec. digit duration condition and .76 correct in the 1-sec. digit duration condition. Longer digit durations improved recall, F (1, 16) $= 10.31$, $p < .01$. Longer interdigit intervals also improved recall, F (1, 16) $= 5.32$, $p < .05$. The *S*s averaged .67 correct in the 0-sec. interdigit interval condition and .75 correct in the 1-sec. interdigit interval condition. Not surprisingly, there was a main effect of retention interval, F (4, 64) $= 134.40$, $p < .001$. Figure 1 presents graphically the significant interaction of digit duration and retention interval, F (4, 64) $= 3.49$, $p < 0.25$. Inspection of Fig. 1 indicates that differences in mean proportion correct associated with the two digit duration conditions were more pronounced at longer than shorter retention intervals. Figure 2 presents the significant Interdigit Interval × Retention Interval interaction, F (4, 64) $= 5.08$, $p < .005$. Inspection of Fig. 2 indicates that differences in accuracy of recall between the two interdigit interval conditions were greater at longer retention intervals. No other main effects or interactions were significant at the .05 level.

EXPERIMENT II

Method

Subjects and design.—Ninety *S*s from the same source as in Exp. I were randomly assigned to the between-*S*s cells of a 2 (digit durations) × 3 (interdigit intervals) × 3 (encoding techniques) × 5 (trials blocks) × 5 (retention intervals) factorial design with repeated measures on the last two factors. Levels of digit duration were again .5 sec. and 1 sec., while levels of interdigit interval were 0, .5, and 1 sec. Three different encoding techniques were employed: no chunking, chunking by two's, chunking by three's. Within each of the five blocks of 10 trials, *S*s were tested twice at each of the same five retention intervals used in Exp. I.

Apparatus and materials.—Each of the 50 six-digit numbers used in this experiment was constructed and recorded in the same fashion as in Exp. I. Again, all .5-sec. digit duration condition tapes involved the additional time compression step in their production. The task interpolated between stimulus presentation and recall was exactly the same as in Exp. I.

Procedure.—The trial structure and procedure for randomizing the order in which retention intervals were tested were the same as in Exp. I. Instructions to *S*s were necessarily different in order that the variable of encoding technique could be manipulated. The *S*s in the no-chunking condition were instructed not to group the digits in any way; they were to store and recall them as a string of six separate digits. The *S*s in the chunking-by-two's condition were told to group them only by two's (e.g., to store 142873 as fourteen, twenty-eight, seventy-three); *S*s in the chunking-by-three's condition were told to group them only by three's (e.g., one hundred forty-two, eight hundred seventy-three). Again, there were six practice trials.

Results

Errors were scored in the same manner as Exp. I. Since each retention interval was tested twice within each 10-trial block, the maximum number of errors for these 2 trials would be 12 (six-digit numbers × two tests). Hence each score entered into the previously described analysis of variance design was the proportion correct of 12.

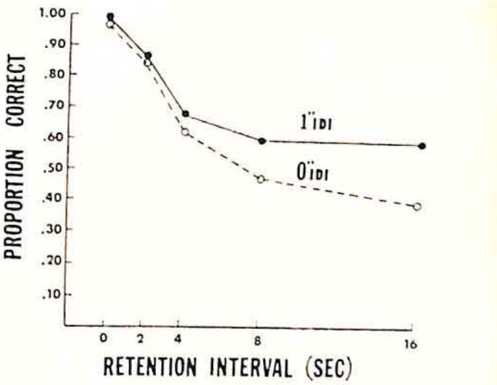


FIG. 2. Interdigit Interval (IDI) × Retention Interval interaction.

The *S*s averaged .47 correct in the .5-sec. digit duration condition and .52 correct in the 1-sec. digit duration condition. The main effect of digit duration was significant, F (1, 72) = 9.82, $p < .005$. Again, longer interdigit intervals improved recall, F (2, 72) = 5.26, $p < .01$. The *S*s averaged .47 correct in the 0-sec. interdigit interval condition, .48 correct at .5 sec., and .54 at 1 sec. A multiple-comparisons procedure (Newman-Keuls throughout this experiment) indicated that all possible pairwise comparisons were significant at the .01 level. Efficient encoding techniques also improved recall, F (2, 72) = 21.34, $p < .001$. Means were .41, .55, and .53 for the no-chunking, chunking-by-two's, and chunking-by-three's conditions, respectively. All pairwise comparisons were significant at the .01 level. Unlike Exp. I, there was a trial-blocks effect, F (4, 288) = 7.68, $p < .001$. Means for the five trial blocks were .49, .47, .50, .53, and .49, respectively. Only the linear component of the trend was significant, F (4, 445) = 8.73, $p < .001$. As in Exp. I, the retention interval effect was significant, F (4, 288) = 426.47, $p < .001$. Analysis of digit duration as a function of retention interval resulted in a significant interaction F (4, 288) = 3.09, $p < .025$. Figure 3 illustrates this interaction. Figure 4 depicts the Interdigit Interval × Retention Interval interaction, F (8, 288) = 3.46, $p < .005$. Figure 5 illustrates the Encoding Technique × Retention Interval inter-

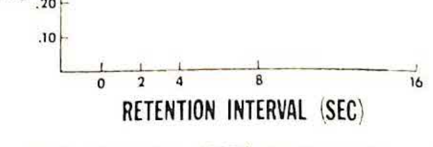


FIG. 3. Digit Duration (DD) × Retention Interval interaction.

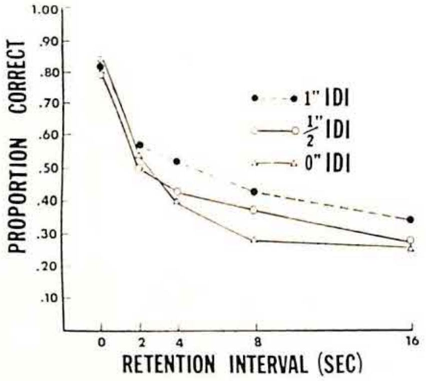


FIG. 4. Interdigit Interval (IDI) × Retention Interval interaction.

action, F (8, 288) = 2.78, $p < .01$. The only other significant interactions were Trial Blocks × Retention Interval, F (16, 1,152) = 5.12, $p < .001$, and Encoding Technique × Trial Blocks × Retention Interval, F (32, 1,152) = 2.10, $p < .005$.

DISCUSSION

As both Exp. I and II found pronounced digit duration and interdigit interval effects, Sitterly's (1968) findings with regard to digit duration and interdigit interval for visual digit sequences can be extended to auditory digit sequences. Apparently, as Haber and Nathanson (1969) also have shown, improvements in recall with decreasing stimulus presentation rate can be accomplished by increases in either digit duration or interdigit interval. The present experiments have not replicated Bergstrom's (1907) findings that increases in stimulus duration do not facilitate recall. The digit duration and interdigit interval effects obtained in both Exp. I and II cannot be explained in terms of increased digit duration and interdigit interval allowing more time for efficient encoding techniques to operate. In Exp. II, efficient encoding techniques (chunking by two's or three's) improved recall across levels of digit duration and interdigit interval and did not interact with either variable. Hence *S*s were apparently able to chunk effectively at even the briefest digit durations and interdigit intervals employed in this experiment. Since increments in digit duration and interdigit interval beyond the briefest intervals resulted in superior recall, *S*s must have been using the extra time for other processing activities, for example, rehearsal.

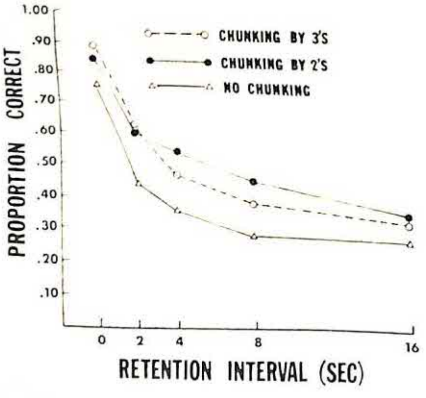


FIG. 5. Encoding Techniques × Retention Interval interaction.

The existence of a trial-blocks effect in Exp. II and not in Exp. I is not surprising. In Exp. II, the positive linear trend for improvement in recall over trials could be attributed to gradual increases in facility with a particular encoding technique. In addition, the four-digit sequences in Exp. I may not have presented enough of a processing load to allow for much increase in the ease of processing over time.

Though the substantial retention interval effects in both Exp. I and II merely replicate the well-known forgetting curves obtained in studies such as that of Peterson and Peterson (1959), the interactions of retention interval with certain other variables are interesting ones. In each of the present experiments, the facilitative effect on recall of longer digit durations and interdigit intervals was greater at longer retention intervals than at shorter retention intervals. Perhaps as suggested above, extra rehearsal time associated with longer digit durations and interdigit intervals made items apprehended under these conditions more resistant to forgetting at longer retention intervals. Certainly increased digit duration and interdigit interval did not differentially influence resistance to forgetting during stimulus presentation; neither did they permit more effective retrieval. An inspection of Fig. 1, 2, 3, and 4 supports these interpretations in that there are no significant differences between digit duration or among interdigit interval conditions at the 0-sec. retention interval. While increasing rehearsal time might make items more resistant to forgetting at longer retention intervals, efficient encoding techniques would seem either to increase resistance to forgetting during stimulus presentation or to permit more effective retrieval regardless of the length of retention interval. The *S*s in the two chunking conditions were presumably able to store each six-digit stimulus in two or three, rather than six storage locations. It is quite possible that retrieval from fewer storage locations is a more error-free process than retrieval from a larger number of these locations. This latter interpretation is supported by the significant differences between each of the two chunking conditions and the no-chunking condition at the 0-sec. retention interval and by the maintenance of these differences as the forgetting curves for the three encoding technique conditions decay at roughly similar rates.

The present experiments, then, have provided at least some concrete indication that errors in short-term recall may be due not only to forgetting caused by inadequate rehearsal, for example, but also to inadequacies in the encoding process. These two sources of error are independently manipulable through variations in encoding instructions and stimulus presentation rate. Further research should allow, under varying experimental conditions, more precise determination of the relative contributions of forgetting and encoding process inadequacies to errors in recall.

REFERENCES

AARONSON, D. Temporal course of perception in an immediate recall task. *Journal of Experimental Psychology*, 1968, **76**, 129–140.

BERGSTROM, J. A. Effects of changes in time variables in memorizing together with some discussion of the techniques of memory experimentation. *American Journal of Psychology*, 1907, **18**, 206–238.

FAIRBANKS, G., EVERITTE, W. L., & JAEGER, R. P. Method for time or frequency compression-expansion of speech. *Transactions of the Institute of Radio Engineers Professional Group on Audio*, AU2, 1954, 7–12.

HABER, R. N., & NATHANSON, L. S. Processing of sequentially presented letters. *Perception & Psychophysics*, 1969, **5**, 359–361.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

MORGANSTEIN, S. Effects of encoding on short-term memory. *Journal of Experimental Psychology*, 1970, **86**, 387–392.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

SITTERLY, T. E. Short-term retention of sequentially presented digits as a function of interdigit interval, digit duration, and series length. *Journal of Experimental Psychology*, 1968, **78**, 174–178.

(Received March 8, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 152–157

NATURE OF THE EFFECT OF VERBAL LABELS ON RECOGNITION MEMORY FOR FORM[1]

TERRY C. DANIEL[2]

University of New Mexico

Four separate sets of form stimuli varying in interval steps along a physically defined shape continuum were developed for use in the experiment. The continua extended from a familiar form (e.g., a duck) at one extreme to an unfamiliar form at the other extreme. The *S*s were presented with the midpoint of one of the form continua, paired with the verbal label which was appropriate for the familiar end of that continuum (e.g., "duck"). The retention test, a single-stimulus recognition memory task employing the midpoint and all other variations from the selected form set, was given immediately, or after a delay of 5 min., 20 min., or 2 days. The introduction of temporal delay produced significant shifts in recognition memory gradients in the direction of the familiar form suggested by the verbal label. Verbal labels may have important effects on the storage stage of memory for form.

Do verbal labels affect the direction of changes in memory for visually presented form stimuli? The answer to this question differs depending upon whether a reproduction or a recognition memory task is employed to assess such changes. It has been shown that learning to associate verbal labels with form stimuli results in superior recognition memory performance (e.g., Ellis, 1968; Ellis & Daniel, 1971) and that, generally, reproduction memory is inferior to recognition memory for form (Rock & Englestein, 1959). The differences between these two memory tasks, however, go beyond a simple comparison of the number of stimuli correctly remembered. Reproduction studies have reported rather consistently evidence for changes in the *qualitative* aspects of *S*s' memory for form stimuli. Drawings of forms have been reported to err in the direction of a more familiar object suggested by a verbal label initially associated with the to-be-remembered form (e.g., Carmichael, Hogan, & Walter, 1932). Recognition studies, even those specifically designed to uncover similar changes (e.g., Prentice, 1954), have characteristically failed to find evidence for such changes.

There are a number of problems involved in regarding this apparent inconsistency in the data as genuine. The reproduction method requires the collection of drawings made from memory by *S*s. As has been previously pointed out (e.g., Riley, 1962), it is virtually impossible to objectively assess the nature or extent of changes that may occur in such drawings. The recognition method requires that a set of stimuli be constructed to serve as alternatives to the target form in the memory test. Furthermore, in order to assess changes in memory, these alternatives must vary systematically along some as-yet undetermined dimension of shape. Problems associated with the construction of suitable sets of alternative forms are largely responsible for the failure of previous attempts to find clear evidence for verbal-label induced changes in recognition memory for form.

In recent years, a number of advances in methodology have suggested a possible means of resolving the inconsistency in findings regarding the effects of verbal labels on memory for form. In particular, a procedure has recently been developed that provides a means of varying a form stimulus along a well-defined dimension of shape qua shape (Ellis & Feuge, 1966; Feuge & Ellis, 1969). By extending this procedure it was possible to investigate memory for visual form under conditions which, while allowing for the observation of the qualitative changes reported in reproduction memory, at the same time made use of the objective response measures of a recognition task. In this instance, the method was applied to the investigation of the phenomenon of "assimilation" that was described by Gibson (1929) and elaborated upon by Carmichael et al. (1932).

---

[1] This paper was taken in part from a doctoral dissertation submitted in partial fulfillment of the requirements for the PhD degree in the Department of Psychology, University of New Mexico. The author wishes to thank Henry C. Ellis, dissertation director, and G. Robert Grice, member of committee, for their guidance and encouragement. This research was supported in part by National Science Foundation Grant, GB-3432.

[2] Requests for reprints should be sent to Terry C. Daniel, now at the Department of Psychology, University of Arizona, Tucson, Arizona 85721.

The experiment specifically tests the hypothesis that a representative verbal label, when associated with an ambiguous visual form stimulus, can affect the course of changes in recognition memory in a manner consistent with that typically found to occur in reproduction memory. Sets of form stimuli varying along a physically defined form continuum were constructed for use in the experiment. The forms were ordered on a directional continuum extending from a highly meaningful, familiar shape at one extreme to a low meaningful, unfamiliar shape at the other extreme. Changes in memory for the target form, the midpoint form on the continuum, were assessed in terms of *S*s' tendency to accept other forms on the continuum as the target form. Disproportionate acceptance of forms differing from the target form in the direction of the more meaningful endpoint, the referent object of the assigned verbal label, would be taken as evidence of a change in memory toward the form suggested by the label.

METHOD

Stimuli.—Simple drawings of common objects—dog, duck, camel, and cat—were selected from a number of children's coloring books. These drawings were selected so as to be of the same general size and area. All of the drawings were transformed by a procedure suggested by Attneave (1954) of fitting the curvilinear drawing with a straight-line polygon. In this manner the drawings were made to conform to solid, straight-lined figures of 24 sides.

The resulting forms (object stimuli) were systematically varied according to general rules suggested by Ellis and Feuge (1966). First a horizontal–vertical axis was constructed at each of the 24 points, the point in each case serving as the origin. For each point on the circumference of the object stimulus a direction of distortion was chosen at random from among the four alternatives; vertical positive and negative (up–down) and horizontal positive and negative (left–right).

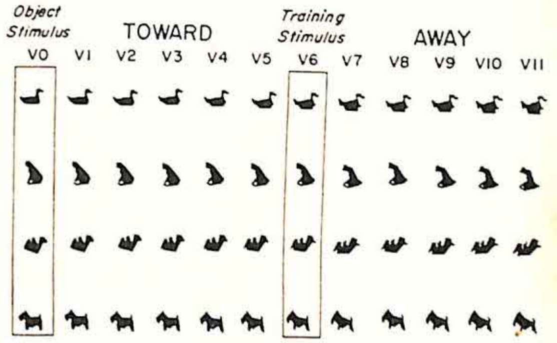


FIG. 1. Object forms and variation stimuli employed in the recognition test.

The first variation (V-1) was constructed by moving each of the 24 points $\frac{1}{20}$ in. (.127 cm.) in the assigned direction and connecting the resulting new set of points in their original order (i.e., Point 1 to Point 2, 2 to 3, etc.) to form a new 24-sided solid figure. The second variation (V-2) was determined by moving the points on the object stimulus again, in the same direction, a distance of $\frac{2}{20}$ in. (.254 cm.). This procedure was followed for the third variation, fourth, etc., until 11 such variations were obtained. The only restriction imposed on the otherwise random assignment of a direction of distortion to each point was that change thus introduced did not destroy the integrity of the figure. In some cases, for example, a change of $\frac{11}{20}$ in. (1.397 cm.) or less in a particular direction for a point might result in cutting the shape into two or more parts. When such an event occurred a new direction was sampled from the three remaining possibilities for that point. This process was continued until an acceptable series of variations was obtained. The four object stimuli and their respective variation sets are presented in Fig. 1.

The series of forms resulting from the variation of an object stimulus represents a continuum of forms spaced at equal intervals along a physically defined shape continuum. Direction along the continuum is defined in terms of the location of the object stimulus or the origin, from which the variations were constructed. Thus, V-6, the training stimulus, is equidistant from V-1 on one end and from the last variation; V-11, on the other.

Apparatus.—All of the forms were photographed and mounted as 35-mm. transparent slides. The stimuli were projected onto a rear-projection screen approximately 20 × 28 cm. by a Kodak Carousel slide projector. The size of the projected image was identical to that of the original figures and thus the extent of variation, e.g., V-1 to V-2, was held constant at $\frac{1}{20}$ in. (.127 cm.) in all cases. Exposure

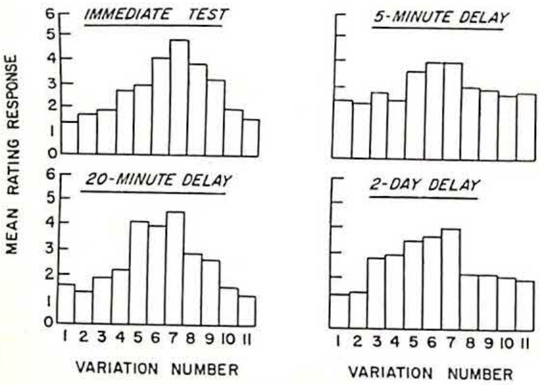


FIG. 2. Gradients of recognition memory at four levels of temporal delay.

durations were controlled by a series of Hunter timers and relays.

Subjects.—A total of 72 *S*s were employed in the study. The *S*s were all college students fulfilling a research participation requirement for an introductory psychology class at the University of New Mexico. They were assigned at random in groups of three to the four experimental conditions.

Experimental design.—The experiment consisted of four independent groups. The *S*s were tested for memory either immediately after exposure to the training stimulus, or after a delay of 5 min., 20 min., or 2 days. Within-*S* variables were direction (toward vs. away from the object stimulus) and distance ($\frac{1}{20}$, $\frac{2}{20}$, $\frac{3}{20}$, $\frac{4}{20}$, and $\frac{5}{20}$ in.) from the training stimulus (V-6). These variables were arranged factorially into a $4 \times 2 \times 5$ analysis of variance design.

Procedure.—The experiment was conducted in two stages. In the first stage (pretraining) *S*s, run in groups of up to three at a time, were exposed to only one form and were instructed that a recognition memory test would follow immediately or after some period of time. Stage 2 of the experiment was a single-stimulus recognition memory test. A detailed description of the procedure follows.

Pretraining.—The *S*s were seated approximately 2 m. from and facing a rear-projection screen. They were instructed that they would be given a brief period of time in which to inspect a form stimulus projected onto the screen and that their task was to remember that form. Complete instructions for the recognition test were given prior to the exposure of the form. The particular object stimulus set (DUCK, DOG, CAMEL, or CAT) presented for any three *S*s was selected at random from among the four sets that had been constructed. After *S*s had been instructed as to the nature of the task, they were told that the form they were to see "resembled a _______," and the form stimulus was exposed for 4.0 sec. In every case, the exposed form was V-6 from the object set selected and the verbal label given was that which was appropriate to the object stimulus of the set (e.g., DUCK).

Recognition test.—The criterion task employed was a single-stimulus recognition test. The test stimuli, consisting of the training form, V-6, and each of the 10 remaining variations from the appropriate object stimulus set were presented 1 at a time in a random order for 4.0 sec. separated by a 5.0-sec. response interval. In accord with a procedure described by Egan (1958), *S*s were required to respond to each form by marking one category of a 6-point scale extending from Positive Yes to Positive No. All *S*s had been given a test sheet with the printed response scale prior to pretraining and had been instructed to use it to reflect their degree of belief that a test form was or was not exactly like the form they had been asked to remember.

The test sheet consisted of four pages of identical printed response scales numbered from 1 to 36. The *S*s were not told how many forms would be presented in the test nor how many actual targets (forms identical to the one they were asked to remember) would be presented. Each of the 11 forms from the appropriate object stimulus set was presented once in the test sequence. The order of stimulus presentation in the recognition test was randomized between subgroups of three *S*s in each delay condition. The only exception to complete randomization of test stimulus order was that no form appeared more than once in the first position at any delay interval.

RESULTS

Rating responses.—For purposes of analysis a numerical value was assigned to each of the degree-of-belief categories. A value of 1 was assigned to the Positive No category, 2 to the Probably No category, etc., with the category Positive Yes being assigned a value of 6. Mean rating responses for each test form at each level of delay are presented graphically in Fig. 2. Although the means for V-6, the training form, are represented it should be noted that only the mean ratings for the new variation forms were included in the analysis. No statistical analysis of the small differences in ratings of V-6 from one delay condition to another was made. The range was from 4.06 at the immediate test to 3.67 after the 2-day delay.

A Delay × Direction × Distance analysis of variance was made of the rating data. Consistent with previous studies of recognition memory for form (e.g., Ellis & Daniel, 1971), the main effect of delay did not reach significance, $F\ (3,\ 68) = 1.00$, $p > .05$. Main effects of both direction and distance of variation reached significance with $F\ (1,\ 68) = 6.32$, $p < .05$, and $F\ (4,\ 272) = 40.77$, $p < .01$, respectively.

Of the two-way interactions, only the Delay × Distance interaction reached significance, $F(12, 272) = 1.86$, $p < .05$.

With regard to the hypotheses discussed in the introduction of this article, the most interesting result of the analysis is represented by the three-way Delay × Direction × Distance interaction. This interaction proved significant, $F(12, 272) = 2.01$, $p < .05$, and the nature of this effect is evident in the graphs of mean rating responses presented in Fig. 2. Inspection of the four histograms reveals that while the overall ratings of new variations did not change as a result of the introduction of delay, the distribution of these responses was altered. There was a clear tendency for *S*s tested immediately after presentation of the training form to give higher ratings to variations that differed from the correct form in a direction away from the object stimulus. In effect, they exhibited a greater preference for forms more distorted from the object stimulus than was the training form they were asked to remember. This tendency was completely reversed by the 2-day delay condition. When *S*s were tested 2 days after presentation of the training form (V-6), they tended to give relatively higher ratings to forms that varied from the training form in the direction of the object stimulus. The preference, then, shifted to forms that were more like the object stimulus.

The 5-min. and 20-min. delay conditions showed an intermediate level of the shift in rating response distribution that was most apparent in the 2-day delay condition.

Signal detection analysis.—The methodology of the "theory of signal detectability" (TSD) has frequently been applied in investigations of recognition memory in recent years. This methodology has been especially useful in single-stimulus recognition tests such as the one employed in this experiment since such tests are quite prone to be affected by response-bias factors. While the observed pattern of shifting rating response gradients is unlikely to have resulted from bias effects alone, the partitioning of sensitivity and bias effects afforded by the TSD analysis could provide additional insight into the origin of the memory shifts.

It was necessary in the present experiment to extend the TSD procedures beyond the typical two-stimulus case, i.e., "old" (correct) and "new" (incorrect). In the present experiment there were in fact 10 qualitatively different new stimuli in the test and it was desirable to distinguish among them in the analysis. Basically this was accomplished by considering each new form (variation) separately with the old training form. Otherwise, the procedure followed for the computation of "hit rates" (HRs) and "false-alarm rates" (FARs) exactly paralleled that described by Hake and Rodwan (1966).

Distance parameters, d', ds, and dm (mean distance of the receiver operating characteristic from the major diagonal) were computed directly from the rating data by a procedure described by G. R. Grice in an unpublished laboratory note.[3] Table 1 presents the distance values computed for each variation in the test at each level of delay. No statistical analysis of differences in these parameters was made, as the sensitivity indices quite clearly reflected the shift in mean ratings depicted in Fig. 2. The effect of 2 days of delay was, in general, to increase FARs for variations toward the object stimulus while decreasing FARs for variations away from the object stimulus. Since the HRs in each case were essentially the same, the result was a shift over delay in the magnitude of the distance metrics. Generally there was a decrease in sensitivity to variations toward the object stimulus and an increase for those away from the object stimulus.

DISCUSSION

The most striking feature of the preceding experiment was the observed effect of temporal delay upon gradients of recognition memory for form. As a result of the presentation of a verbal label and the introduction of delays of up to 2 days, the distribution of rating

---

[3] GRICE, G. R. Computations in the use of TSD. Unpublished laboratory note, 1968. Copies of this laboratory note are available from T. C. Daniel.

TABLE 1

Computed *d*′, *d*s, and *d*m Values for all Variations at Each Level of Delay

Delay cond.	Stimulus variation no.									
	1	2	3	4	5	7	8	9	10	11
Immediate test										
d′	.88	2.93	1.60	.83	.69	−.48	.08	.51	1.26	.84
*d*s	1.39	1.71	1.42	.78	.70	−.48	.08	.50	1.27	1.33
*d*m	2.41	1.39	1.34	.77	.70	−.49	.09	.50	1.28	2.33
5-min. delay										
d′	2.04	1.55	1.45	1.60	.51	.65	1.10	1.28	1.21	2.38
*d*s	1.33	1.48	1.16	1.50	.58	.72	1.26	1.17	1.38	.96
*d*m	1.40	1.48	1.23	1.51	.42	.62	1.20	1.20	1.34	1.21
20-min. delay										
d′	2.43	1.40	5.49	1.22	−.09	−.48	.77	.90	2.37	1.69
*d*s	1.74	2.03	1.73	1.13	−.09	−.46	.61	.75	1.73	2.45
*d*m	1.46	3.27	1.23	1.09	−.13	−.42	.59	.72	1.47	4.05
2-day delay										
d′	.76	.74	.60	.40	.12	−.08	.52	.52	.54	1.58
*d*s	1.16	1.13	.55	.45	.09	−.10	.82	.82	.85	1.26
*d*m	2.22	2.17	.52	.50	.11	−.20	1.67	1.64	1.72	1.08

responses showed a pronounced shift toward a more meaningful form, in this case represented by the object-stimulus referent of the verbal label. This observed shift in the distribution of false recognitions occurred even though the overall level of performance, measured, either in terms of responses to the correct form or of the overall level of responses to incorrect forms, was unchanged. These data are consistent with findings that have characteristically been confined to reproduction memory studies (e.g., Carmichael et al., 1932).

While these results principally point to the role of verbal labels in memory for form stimuli, it should be noted that the introduction of the label very likely influenced *S*s' initial encoding of the form as well. Even without the explicit presentation of the verbal label, there is ample evidence (e.g., Ellis & Homan, 1968) that *S*s would have provided their own labels to the form stimuli, labels which would no doubt have been very similar to those provided by *E*. The focus of the present experiment was, however, upon the effect of the label on the storage stage of memory for the form with the influence of both encoding and retrieval factors minimized. Thus the observed shifts in recognition memory gradients provide evidence of changes taking place during the storage stage of memory, particularly during what is generally termed long-term memory.

Somewhat contrary to these findings, other recent investigations (Daniel & Ellis, 1972; Ellis & Daniel, 1971) have tended to emphasize the stability of long-term recognition memory for form stimuli. There are, however, several important differences in the procedures employed in these previous studies and that used in the current one. Both of the earlier experiments made use of a multiple-choice type of recognition test requiring *S*s to select the originally learned shape from a set of distractors; the principal focus being on *S*s' ability to select the correct form. The present study employed a single-stimulus test that required *S*s to judge independently whether each shape was or was not the target form. This procedure allowed for assessment of *S*s' tendency to accept specific forms, primarily centering upon *S*s' tendency to accept incorrect forms. More important, the experiments differ with regard to the nature of the distractors used in the recognition test. Both of the previous studies used random variations of the target forms as distractors. The experiment reported here used distractors which varied systematically from the target form, specifically in terms of their similarity to the referent object of the assigned verbal label.

If the results of the Daniel and Ellis (1972) and Ellis and Daniel (1971) experiments are considered along with the findings of the present study, several conclusions may be drawn about the course of recognition memory for form. First, all three experiments indicate that recognition memory, measured in terms of *Ss'* tendency to select the target form, is quite stable over a rather extensive range of retention intervals. Analysis of the pattern of responses to incorrect forms, provided by the present experiment, revealed that important changes in memory are nonetheless occurring during the retention interval. These changes may be interpreted as shifts in the memory strength of forms similar to the target form. Specifically, alternative forms which differ from the target in the direction of being more similar to a form suggested by the verbal label associated with that target tend to gain in strength, while forms varying away from the form referred to by the label tend to lose in strength. The two previous studies could not have detected such changes as the use of random variations would make it highly unlikely that an alternative form would fit the assigned label better than the original form.

REFERENCES

ATTNEAVE, F. Some informational aspects of visual perception. *Psychological Review*, 1954, **61**, 183–193.

CARMICHAEL, L., HOGAN, H. P., & WALTER, A. A. An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology*, 1932, **15**, 73–86.

DANIEL, T. C., & ELLIS, H. C. Stimulus codability and long-term recognition memory for visual form. *Journal of Experimental Psychology*, 1972, **93**, 83–89.

EGAN, J. P. Recognition memory and the operating characteristic. (Tech. Rep. No. AFCRC-TN-58-51) Bloomington: Indiana University, Hearing and Communication Laboratory, 1958. (DDC-AD 152 650)

ELLIS, H. C. Transfer of stimulus predifferentiation to shape recognition and identification learning: The role of properties of verbal labels. *Journal of Experimental Psychology*, 1968, **78**, 401–409.

ELLIS, H. C., & DANIEL, T. C. Verbal processes in long-term stimulus recognition memory. *Journal of Experimental Psychology*, 1971, **90**, 18–26.

ELLIS, H. C., & FEUGE, R. L. Transfer of predifferentiation training to gradients of generalization in shape recognition. *Journal of Experimental Psychology*, 1966, **71**, 539–542.

ELLIS, H. C., & HOMAN, L. E. Implicit verbal responses and the transfer of stimulus predifferentiation. *Journal of Experimental Psychology*, 1968, **76**, 486–489.

FEUGE, R. L., & ELLIS, H. C. Generalization gradients in recognition memory of visual form: The role of stimulus meaning. *Journal of Experimental Psychology*, 1969, **79**, 288–294.

GIBSON, J. J. The reproduction of visually perceived forms. *Journal of Experimental Psychology*, 1929, **12**, 1–39.

HAKE, H. W., & RODWAN, A. S. Perception and recognition. In J. Sidowski (Ed.), *Experimental methods and instrumentation in psychology*. New York: McGraw-Hill, 1966.

PRENTICE, W. C. H. Visual recognition of verbally labeled figures. *American Journal of Psychology*, 1954, **67**, 315–320.

RILEY, D. A. Memory for form. In L. Postman (Ed.), *Psychology in the making*. New York: Knopf, 1962.

ROCK, I., & ENGELSTEIN, P. A study of memory for visual form. *American Journal of Psychology*, 1959, **72**, 221–229.

(Received March 13, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 158–163

MEMORY SEARCH FOR CVC AND CCC TRIGRAMS[1]

STEPHEN F. CHECKOSKY[2] AND NORIK BABOORIAN

Lake Forest College

The *S*s participated in a series of memory search tasks in which the memory sets were comprised of one or two CVC or CCC trigrams. In half of the conditions the test stimuli were trigrams, and *S*s were asked to indicate whether that trigram matched any trigram in the memory set. In the remaining conditions the test stimuli were single letters, and *S*s were asked to indicate whether that letter matched any letter in the memory set trigram(s). In general, *S*s were faster in matching trigrams. This effect was independent of memory-set size but interacted strongly with response type (YES vs. NO). The *S*s were also faster when the memory sets were comprised of CVC trigrams. There is no interaction between this and any of the other experimental factors. Finally, *S*s generally were slower when the type of comparison which they accomplished (letter or trigram) varied randomly over a series of trials. This effect was also independent of memory-set size but interacted with response type and comparison type. These data were interpreted in terms of information-processing stages according to S. Sternberg's additive factor method.

Recent theories of information processing in recognition tasks have postulated a sequence of stages. Sternberg (1969) has offered both a methodology for investigating information-processing stages and a theoretical description of such stages in one type of recognition experiment, the memory search experiment. In a memory search experiment, a memory set consisting of a short list of stimuli is learned prior to the presentation of a test stimulus. The *S*'s task is to decide whether or not the test stimulus is a member of the memory set and to make an appropriate response.

The methodology (additive factor method) proposed by Sternberg (1969) presumes that total reaction time (RT) in recognition tasks is composed of the sum of the times to complete a sequence of information-processing stages. The additive factor method can be conceived of as a technique for isolating the components of total RT. The basic notion is that when two or more factors are manipulated in a recognition task, if they are mediated at a single information-processing stage, their joint effect on mean RT would very likely be some sort of interaction. If they are mediated at separate independent stages, their joint effect on mean RT should be additive.

The theory of processing in a memory search task offered by Sternberg (1969) is cast in the form of a description of operations at various information-processing stages. The stages so described were isolated in a series of memory search experiments which were analyzed and interpreted according to the additive factor method. Central to this theory is a contention that the increase in RT for each unit increase in the number of items in the memory set (*M*) reflects the time required to compare a representation of the test stimulus with an item in memory. A puzzling feature of these comparison operations is that they are presumed not to terminate when a match is encountered. Instead, *S* continues his comparison operations until the test stimulus has been compared with each memory-set item. This "exhaustive" processing of the memory-set items is inferred from the independence of the effects of *M* and response type. If comparisons terminated upon the detection of a match, the effect of *M* would be half as large for YES than for NO responses.

In order to account for the evidence suggesting exhaustive processing of the memory-set items, Sternberg (1969) proposed

---

[1] The present research was supported by Research Grant MH 18702-01 and MH 19863-01 from the National Institute of Mental Health.

[2] Requests for reprints should be sent to Stephen F. Checkosky, Department of Psychology, Lake Forest College, Lake Forest, Illinois 60045.

that *S* receives no *immediate* feedback as to the outcome of each individual comparison. In order to evaluate whether or not a match has been encountered an additional operation (register examination) is required. However, switching from comparison to register-examination operations takes considerably longer than the time required to complete an additional comparison operation. Thus an economy is provided by minimizing the number of times register examination takes place per trial. This can be accomplished by completing all *M* comparisons, followed by a single register examination.

The major purpose of the present experiment was to provide a test of Sternberg's (1969) exhaustive comparison and register-examination theory. The test was provided by varying *M* along with another factor which is presumed to influence comparison time. This factor is the type of comparison that *S* was asked to accomplish (C). The memory sets in the present experiment were comprised of letter trigrams. The two comparison types were trigram comparison and letter comparison. In the trigram comparison conditions the test stimulus was a trigram, and *S* was asked to respond YES if that trigram matched *any trigram* in the memory set. In the letter comparison conditions the test stimulus was a single letter, and *S* was asked to respond YES if that letter matched a letter in any of the memory-set trigrams. There were either 1 or 2 trigrams in the memory set. The expected source of the C × *M* interaction can be derived quite simply. Should an effect of the C factor emerge, and be attributable to comparison operations, then the magnitude of this effect would be greater when $M = 2$ (2 comparisons) than when $M = 1$ (1 comparison).

A second purpose of the present experiment was to "localize" the effect of pronounceability (P) of memory set and test stimuli. Egeth and Blecker (1971) using a same-different matching task found that identical CVC trigrams low in meaningfulness (e.g., BIF) were matched considerably (168 msec.) faster than low meaningful CCV trigrams (e.g., BFI). In half of the conditions of the present experiment, the memory sets were comprised of CVC trigrams. In the remaining conditions, the memory sets were comprised of CCC trigrams. Test stimulus trigrams in the CVC conditions were all CVCs. Test stimulus trigrams in the CCC conditions were all CCCs. If an effect of pronounceability were exclusively due to the time to form a memorial representation of (encode) the test stimulus, then there should be no P effect when the test stimulus was a single letter. If P influenced the time to process each memory set item, then the magnitude of the P effect would be greater at $M = 2$ than $M = 1$.

A third purpose of the present experiment was to localize the source of any adjustments to variations in comparison type expectancy (E). In half of the conditions (pure conditions), the comparison type (trigram or letter) remained constant through a series of trials. Thus *S* knew prior to each trial what type of comparison was required on that trial. In the remaining conditions (mixed conditions) the type of comparison was varied randomly over a series of trials. Thus *S* could not know what type of comparison was required until the test stimulus appeared (and the RT clock had begun). Again in accord with the additive factor method, localization of the E effect would depend on the interactions between E and the remaining experimental factors.

METHOD

Subjects.—Twelve right-handed female undergraduates of Lake Forest College volunteered to participate as *Ss*. These *Ss* were paid for their services.

Apparatus.—Stimuli were presented in three in-line displays (Lehigh Valley Standard Pattern No. 700, 701, and 702) which were arranged horizontally. Each display was circular with a diameter of 1 in. The horizontal distance between the center of a pair of displays was $1\frac{3}{4}$ in. Stimuli were uppercase white letters which were constructed for lines of $\frac{1}{8}$ in. thickness. The preexposure field, postexposure field, and the background field were all a uniform black. The displays were set in a 3 × 3 ft. wooden frame which was painted brown. The *S*'s index fingers rested on metal keys which activated a microswitch.

Reaction times were measured on a Beckman Universal EPUT and Timer (Model 5230) which was set to be accurate to the nearest $\frac{1}{100}$ sec. A set of lights on *E*'s panel indicated *S*'s response. The sequence of ready signal, presentation, and reset was controlled by a Lafayette Instrument Co. eight bank timer (Model 5431A). Errors and RT were manually recorded.

Design.—Each *S* participated in the experiment for 6 days. On each day, *S* would receive six blocks of trials. Four blocks (24 trials each) resulted from a combination of two values of memory-set size (one vs. two trigrams) and two comparison types (single-letter match vs. trigram match). In the two remaining blocks (48 trials each), the type of comparison required was varied randomly over trials. In one of these blocks $M = 1$ trigram, and in the other $M = 2$. Six practice trials preceded each block. The same trigram would comprise the memory set at all three $M = 1$ conditions on a given day. The same pair of trigrams would comprise the memory set at all three $M = 2$ conditions on a given day. On alternate days, *S*'s memory sets would be comprised of CVC and CCC trigrams. The order of presentation of the six blocks of trials was counterbalanced for each *S* over the 6 days. Further, this order was counterbalanced over the 12 *S*s on each day. On half of the trials in each block a YES response was correct; on the remaining trials a NO response was correct. The sequence of trials was arranged so that the probability that a test stimulus was a member of the memory set on Trial $n - 1$ was independent of its membership on Trial *n*.

Stimuli.—The CVC trigrams employed as memory-set members in this experiment were RIC, JUB, and FET ($M = 1$ conditions), also KAB and POM, FAS and BIP, KOD and TUP ($M = 2$ conditions). The CCC trigrams were RMC, JRB, and FWT ($M = 1$ conditions), also KHB and PLM, FHS and BMP, KLD and TRP ($M = 2$ conditions). Each of these trigrams was a memory-set member for each of the *S*s on one experimental day.[3]

Test stimuli in the trigram-match conditions which required a NO response always had two letters in common with one memory-set trigram. The relative position of these letters on the screen corresponded to their relative position in the memory-set trigram. For example, if the memory set was comprised of the trigrams KAB and POM, a test stimulus such as FOM might appear, but never MOF (letter positions do not correspond) or KOM (letters shared with more than one memory-set trigram). In addition, when the memory set was comprised of CVCs, all test stimuli in the trigram-match conditions were CVCs. When the memory set was comprised of CCCs, all test stimuli in the trigram-match conditions were CCCs.

Test stimuli in the letter-match conditions which required a YES response all appeared in the same relative position on the display as in the memory-set trigram. For example, if KAB and POM comprised the memory set and K appeared as a test stimulus, it would appear in the left inline display; M would appear on the right. When the memory set was comprised of CVCs all letters which appeared in the middle screen were vowels. When CCCs comprised the memory set, all letters which appeared in the middle screen were consonants. On exactly one-third of the trials in each block in the letter-match condition, the test stimulus would appear in the left display. On exactly one-half of these trials the test stimulus would match a memory-set letter. The same was true of the center and right displays. In the trigram-match conditions, when a NO response was required, on exactly one-third of the trials it was the letter on the left display which failed to match. On one-third of the trials it was the letter in the center, and on one-third of the trials it was the letter on the right.

Procedure.—The *S* was seated directly in front of the display with her index fingers resting lightly on the metal keys. The trigram(s) which comprised a memory set were printed on an index card by *E* and placed in front of *S* while instructions as to S–R assignments were verbally given. Half of the *S*s were instructed to respond YES by depressing the right-hand switch; for the remaining *S*s these instructions were reversed. The *S*s were encouraged to strive more for accuracy than speed. After each correct response, *S* was told her RT to the nearest $\frac{1}{100}$ sec.

RESULTS

Days 1 and 2 were considered as practice in order to allow *S*s to become more familiar with the tasks and to give time for RTs to stabilize. The mean correct reaction time for each of the experimental conditions was computed for Days 3–6. An analysis of variance was performed on these data. The design of this analysis was a $2 \times 2 \times 2 \times 2 \times 2$ factorial with all of the factors within *S*. The five factors in the present experiment were: comparison type (C), trigram versus letter; memory-set size (*M*), 1 versus 2 trigrams; pronounceability (P), CVC versus CCC; comparison type expectancy (E), pure versus mixed; re-

---

[3] The stimuli in the present experiment were not counterbalanced over the $M = 1$ and $M = 2$ conditions. In order to evaluate whether the stimuli in these conditions were equal in difficulty, a separate experiment was conducted. Twelve *S*s received each of the nine CVC and CCC stimuli in $M = 1$ trigram-match conditions (24 trials/block) on 2 successive days in counterbalanced order. The results of this experiment showed no significant difference in RT between stimuli employed in the $M = 1$ conditions of the main experiment (480 msec.) and those employed in the $M = 2$ conditions (484 msec.).

TABLE 1

Mean Correct Reaction Time (RT) (in msec.) and Percent Errors (%E)

Comparison and response	M	CVC trigrams in memory				CCC trigrams in memory			
		Pure condition		Mixed condition		Pure condition		Mixed condition	
		RT	%E	RT	%E	RT	%E	RT	%E
Trigram									
YES	1	372	2.8	394	.7	396	1.4	415	1.4
	2	439	8.3	453	5.3	462	4.2	469	5.9
NO	1	420	2.8	477	5.2	445	3.5	495	3.5
	2	501	6.2	524	5.2	524	3.8	548	9.7
Letter									
YES	1	412	3.5	438	5.2	425	3.8	446	5.2
	2	478	8.3	486	9.7	478	13.5	514	11.5
NO	1	460	3.5	456	2.4	470	3.1	469	2.8
	2	518	4.5	518	4.2	533	5.6	538	5.6

Note.—M = number of items in the memory set.

sponse type (R), YES versus NO. The percentage of times that *S* made an incorrect response was also calculated at each level of the above factors. Table 1 presents mean RT and error percentage for each combination of the experimental factors.

When the memory set was comprised of CVC trigrams *S*s were 17 msec. faster than when the memory set was comprised of CCC trigrams. This difference was significant, $F\ (1,\ 11) = 12.1$, $p < .01$. The magnitude of the P effect was greater when the test stimulus was a trigram (22 msec.) than when the test stimulus was a single letter (13 msec.). The P × C interaction, though large relative to the main effect of P, was not significant, $F\ (1,\ 11) = 2.0$, $p > .10$. A t test was then performed in order to determine whether there was a significant P effect when the test stimulus was a single letter. The resulting $t\ (11)$ was 2.18, $p < .05$, one-tailed. The $P \times M$ interaction was clearly nonsignificant, $F\ (1,\ 11) = .3$. The effect of P when $M = 1$ was 16 msec.; when $M = 2$ the effect was 19 msec. No other interactions in the present experiment involving P were significant.

Evidence was obtained that the memory-set trigrams were exhaustively processed. The effect of M for YES responses was 60 msec., for NO responses, 64 msec. The $M \times R$ interaction was not significant, $F\ (1,\ 11) = .9$. The overall effect of M (62 msec.) was highly significant, $F\ (1,\ 11) = 256.1$, $p < .001$, as was the overall effect of R (51 msec.), $F\ (1,\ 11) = 69.7$, $p < .001$.

In general, *S*s were faster in the trigram comparison conditions. This difference (19 msec.) was significant, $F\ (1,\ 11) = 10.2$, $p < .01$. There was one reversal of this effect in the present experiment, in the mixed condition NO responses. Here the trigram comparisons averaged 16 msec. slower than the letter comparisons. This reversal produced the following significant interactions: $C \times R$, $F\ (1,\ 11) = 59.1$, $p < .001$; $C \times E$, $F\ (1,\ 11) = 8.8$, $p < .025$; and $C \times R \times E$, $F\ (1,\ 11) = 18.3$, $p < .005$.

The overall effect of C when the memory set was comprised of a single trigram was 20 msec. When the memory set was comprised of two trigrams the C effect was 18 msec. The $C \times M$ interaction was not significant, $F\ (1,\ 11) = .1$. There was, however, a singificant $C \times M \times E$ interaction, $F\ (1,\ 11) = 11.1$, $p < .01$. The effect of C for NO responses in both the pure and mixed conditions was *less* at $M = 2$ than at $M = 1$. The interaction resulted since the C effect was in opposite directions at these conditions.

DISCUSSION

Before addressing theories of the M effect, the locus of the C effect needs to be evaluated. Operationally, what is varied along with instructions to S is the number of elements in the display. This variable might be expected to influence only stimulus input operations. However there are two aspects of the present data which suggest that the obtained effect of C reflects comparison rather than stimulus input operations. First the trigram conditions were generally faster than the single-letter conditions. If the effect of C were exclusively at the stimulus input stage, the opposite result would be expected. Second, this factor produced a large and significant interaction with response type. Again this would not be expected if C influenced only stimulus input time.

Given the evidence that the obtained effect of C reflects comparison operations, then the joint effect of C and other experimental factors can be evaluated in order to elucidate the nature of the operations at the comparison stage. There are two sources of data in the present experiment that are seriously inconsistent with Sternberg's (1969) exhaustive comparison and register checking theory. First, this theory would predict that the obtained effect of C would be greater at $M = 2$ than at $M = 1$. This did not obtain. Second, since S is presumed to wait until all comparisons are completed before selecting a response, no C × Response Type interaction would be predicted. Yet this interaction was large and significant. Further, in the mixed conditions, the effect of C was in opposite directions for YES and NO responses. It is difficult to imagine how any variation of an exhaustive comparison and register checking theory of the M effect could account for these data.

If M does not reflect comparison time what does it reflect? Checkosky (1971) has argued that M reflects the time to generate a visual representation for each memory-set item. According to this view, generation represents a type of constructive operation which recent "outflow" theories (Kolers, 1968; Neisser, 1967) have postulated as being central to the accomplishment of recognition. Comparison of the test stimulus with these visual representations occurs subsequent to generation and is characterized by self-terminating feature testing. Exhaustive processing of the memory-set items follows directly from this conception since comparison and hence response selection does not begin until generation is completed. Thus, in order to minimize errors, S would have to generate a representation of all of the memory-set items. The overall independence of M and C would be expected since these factors influence separate stages. Finally, since comparison operations are presumed to be self-terminating, the C × R interaction found in the present experiment is consistent with this theory (though not explicitly predicted). The major departure of the present data from the expectations of generation theory is reflected in the significant C × M × E interaction. This implies that the effect of M was not consistent over the four conditions resulting from each combination of C and E. The effect of M was greatest in the pure trigram conditions (73 msec.) and least in the mixed trigram conditions (59 msec.). The processes responsible for these data are unclear.

The second purpose of the present experiment was to "locate" the P effect in an information-processing sequence by analyzing the data in accordance with the additive factor method. One possibility is that P influences the time to form a memorial representation of (encode) the test stimulus. However, the expected P × C interaction did not emerge. Further there was a significant effect of P when the test stimulus was a single letter. From these data it can be concluded that the stimulus input stage is not the exclusive locus of the P effect. Further it remains unclear whether there is any P effect at this stage.

The independence of M and P suggests that memory processing time for trigrams in a memory search task does not depend on whether these trigrams can be readily organized into pronounceable units. This provides strong support for Sternberg's (1969) contention that the functional memorial representation of the memory-set items in a memory search task is not the name (i.e., auditory representation) of these items but instead is some sort of visual representation.

A third purpose of the present experiment was to evaluate the effect E. The magnitude of the E effect varied widely across levels of C and R. The resulting interactions argue strongly that E influences operations at the comparison stage. A closer examination of the data reveals that E produced its greatest effect (39 msec.) for the NO responses in the trigram comparison conditions. One admittedly ad hoc approach to an account for these data would focus on interference provided by the two letters in the test stimulus trigram which match letters in the memory-set trigram(s). Perhaps the presence of matching letters in the test stimulus impairs performance most when there is prior uncertainty as to whether detecting such a match should result in a YES response.

At the same time, E produced no effect for the NO responses in the letter-comparison condition. This absence of an E effect at a condition defined by a combination of C and R argues strongly that there is *no* adjustment to variations in E at any stage *prior* to the comparison stage. If there were such an adjustment, it would have to manifest itself by an E effect at all levels of C and R.

REFERENCES

CHECKOSKY, S. Speeded classification of multidimensional stimuli. *Journal of Experimental Psychology*, 1971, **87**, 383–388.

EGETH, H., & BLECKER, D. Differential effects of familiarity on judgments of sameness and difference. *Perception & Psychophysics*, 1971, 9, 321–326.

KOLERS, P. A. Some psychological aspects of pattern recognition. In P. A. Kolers & M. Eden (Eds.), *Recognizing patterns*. Boston: MIT Press, 1968.

NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

STERNBERG, S. The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 1969, **30**, 276–315.

(Received March 17, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 164–169

SUBVOCAL ACTIVITY AND ACOUSTIC CONFUSIONS IN SHORT-TERM MEMORY[1]

WILLIAM E. GLASSMAN[2]

University of Toronto, Ontario, Canada

Research on the source of acoustic confusions in short-term memory has emphasized possible auditory and articulatory factors. The present study examines the role of subvocal activity as the basis of an articulatory (speech-motor) mechanism. Forty-eight females served in an experiment manipulating degree of subvocalization (Minimizing) by a self-conditioning procedure; degree of acoustic similarity of items was also varied, using a proactive-inhibition-release paradigm. Analysis of overall recall showed significant effects of Similarity and Trials × Similarity. Error analyses indicated Minimizing affected the proportion of acoustic errors, while Similarity did not. The results were interpreted as suggesting that subvocal rehearsal produces acoustic confusions, but that minimizing activity may lead to encoding in other ways.

Since the initial discovery by Conrad (1964) that patterns of errors in short-term memory (STM) resembled those in the perception of speech, there has been considerable interest in the source of these "acoustic" confusions.[3] Much of the debate has centered on the relative importance of articulatory (Hintzman, 1967) and auditory (Sperling & Speelman, 1970) factors. According to the former hypothesis, the errors are due to items being stored as phonemes in an articulatory system, probably based on speech-motor activity. The auditory hypothesis attributes the errors to confusion of allophones in an auditory storage system (in the case of visual presentation, there is an implied recoding process). As Wickelgren (1969) has noted, these hypotheses are not easily discriminable operationally, and previous arguments based on evidence from confusion matrices have been inconclusive.

Despite this seeming impasse, there has been considerable research on the role of articulation in producing acoustic confusions. Murray (1966), varying degree of vocalization at presentation, found that greater vocalization facilitated overall recall but that as vocalization increased there was a corresponding increase in the proportion of acoustic errors. Later studies of articulation have generally involved introduction of a competing response, a technique developed by Murray (1967). Levy (1971) and Peterson and Johnson (1971) have used this method to minimize articulation at presentation. As Hintzman (1967) has noted, the level at which articulatory factors operate could either be central or involve peripheral (subvocal) speech-motor responses. The fact that the competing response method generally produces a recall decrement (e.g., Peterson & Johnson, 1971) argues for either a peripheral process, or possibly a central process which cannot go on independently of competing speech activity.

In a study using suppression (tongue held between teeth), Gumenik (1969) reported no difference in overall recall for restricted articulation as compared to silent reading. However, it is unclear whether there was a change in number of confusion errors between these conditions. This point, plus the possibility that the suppression technique was ineffective in

---

[1] This research was supported by Grant APA 146 from the National Research Council of Canada and OMHF 164 from the Ontario Mental Health Foundation to Bennet B. Murdock, Jr., and by a National Research Council Postgraduate Fellowship to the author. The author wishes to express his appreciation to Bennet B. Murdock, Jr., for his guidance during the execution and reporting of this research, and to John C. Ogilvie for his advice on statistical matters.

[2] Requests for reprints should be sent to William E. Glassman, Department of Psychology, University of Toronto, Toronto 181, Ontario, Canada.

[3] The term is used hereafter to refer to the phenomenon, without implying that the errors are acoustic in origin.

inhibiting the *relevant* activity, makes interpretation of these data difficult.

While the evidence suggests that vocal activity plays a role in the production of acoustic confusions, previous studies have not clearly established the nature of the underlying mechanism. With regard to the articulatory hypothesis of Hintzman (1967), the effects of a competing response may be interpreted as either suppressing formation of a peripheral encoding or interfering with an as yet undefined central process. Conrad (1964) had originally suggested that acoustic errors might be due to rehearsal which is mediated by subvocalization, but undertook no effective test. Subvocalization consists of attenuated movements of the laryngeal muscles, and as such represents a form of peripheral speech-motor activity. Since articulation clearly depends upon subvocal activity, all prior studies involving articulation of necessity also involved subvocalization, and one cannot determine from them the relative importance of the subvocal and overt responses. Given this, it is desirable to examine the role of subvocal activity in the absence of overt responses.

The meaning of subvocal activity, in terms of an articulatory hypothesis (Hintzman, 1967), has not been clearly established. Most investigators have assumed that subvocalization is a form of speech (see review by Locke, 1970), which would imply its utility as a rehearsal mode. The problem has been that while overt speech activity (articulation, vocalization) must be accompanied by subvocal activity, there is no logical necessity that speech-muscle activity in the absence of an overt response need represent a form of speech at all. However, recent evidence by Locke and Fehr (1970) supports the notion of subvocal rehearsal as a form of speech. Given this, one can formulate hypotheses about the relation of subvocalization to the production of acoustic confusions in STM. Four possibilities seem to exist: (*a*) If there is no meaningful relation, one would expect no effect on performance of manipulating subvocalization. (*b*) If subvocalization mediates all STM, one would expect a drop in overall recall when subvocalization is minimized. This would seem the suitable prediction on the basis of studies of degree of overt vocalization (Murray, 1966). (*c*) If subvocalization mediates all forms of *encoding*, but acoustic confusions occur only when it is used as a form of *storage*, then one might expect that minimizing subvocal activity would lead not only to a drop in overall recall (less efficient encoding) but also to a drop in proportion of acoustic errors (lower probability of subvocal storage). (*d*) If subvocalization represents the source of acoustic confusions, but is only one of several forms of storage in STM, then one would predict no drop in overall recall but a drop in the proportion of acoustic errors when subvocalization is minimized. This hypothesis would seem to be in accord with Shulman's (1971) notion of plasticity in STM.

In order to test these hypotheses, one must measure subvocal activity without interfering with it. The only effective way to do this is by direct monitoring of speech muscle activity by electromyogram (EMG). It is desirable both to measure the normal pattern of activity and to vary it in accord with the test of the predictions. The present technique allows both measurement of ongoing activity and also manipulation of amount of activity by providing feedback for self-conditioning; hence, it allows a direct attack on the role of subvocalization as the source of acoustic errors. Since the focus is not on the possible role of auditory factors, visual presentation was used. Acoustic similarity of items was varied; in the high-similarity groups, items on paired trials were maximally similar, providing high interference, with "proactive-inhibition (PI) release" (Wickens, 1970) possible between pairs. Hence the experiment provides an examination of the relative roles of subvocal activity and acoustic similarity in the production of acoustic confusions.

METHOD

Subjects.—Forty-eight females of college age served as *Ss*. Some were paid volunteers, while

TABLE 1

ITEMS USED AS RETENTION TASK IN EXPERIMENT

High similarity	Low similarity
Cat-Goal-Nab	Rig-Hoe-Melt
Cab-Bowl-Nap	Sponge-Dash-Flee
Dale-Dope-Taught	Rope-Pit-Cute
Pail-Dupe-Sought	Test-Kin-Plum
Mad-Bit-Slap	Male-Slide-Reed
Mat-Kit-Slab	Swamp-Elm-Raw

others participated as part of the requirements for psychology courses at the University of Toronto.

Design.—The *S*s were either given training to minimize subvocalization or were simply told that their subvocal activity would be monitored. For the test, a Brown-Peterson paradigm was used, with word triads as test items. All triads were balanced for Thorndike and Lorge frequency of occurrence. In the low-similarity conditions, items were randomly related; in the high-similarity conditions, items on paired trials were maximally acoustically similar, with PI release possible between pairs (i.e., A, A′, B, B′, C, C′). All of the *S*s received a practice trial of digits and six test trials. Half of the *S*s in each subvocalization condition received high-similarity items, while half received low-similarity items. Each of six *S*s in a cell received 1 of the 24 possible forward permutations of item order, randomly chosen. The whole design was replicated, yielding a 2 × 2 × 2 × 6 × 6 factorial (Similarity × Minimizing × Replications × *S*s × Trials). (See Table 1 for items.)

Apparatus.—Subvocal activity was recorded by 9-mm. silver disc electrodes placed lateral to the voice box, with ground to either right earlobe or back of neck. (The technique is a modification of that used by Hardyck, Petrinovich, & Ellsworth, 1966.) Amplification was through a Grass Model 7 polygraph, using a Model 7P5 wide-band electroencephalogram (EEG) integrator preamplifier, integrated over a .015-sec. time constant. Output was recorded on a chart recorder driven at 2.5 mm/sec, with pen deflection representing 30 μv/cm. Feedback was provided by a buzzer connected to the polygraph output through a low-hysteresis Schmitt trigger; the triggering level was variable by *E* and set to match a visual criterion of base-line pen deflection (i.e., average amplitude for *S* at rest).

Items for the memory task were projected using a Shibaden closed-circuit TV, which monitored items presented by an IBM Selectric typewriter; the typewriter carriage had been masked for single-item presentation. A solenoid was used to advance the carriage for successive presentations and was controlled by a Diapilot that monitored blips prerecorded on a Sony 250 tape recorder. This procedure allowed control of all presentation and intertrial intervals. Initial recording of the tape was done with the aid of several Hunter 100C timers.

Procedure.—The *S*s in both Minimizing conditions were told that the experiment was concerned with studying "the relationship of speech activity to various forms of thinking tasks." After an accommodation period to get used to the apparatus, all of the *S*s received preliminary tasks (reading, free recall lists) designed to provide base-line information on each *S*'s subvocalizing; the tasks were presented to *S*s as part of the experiment. Those in the control groups went through all tasks without *E* commenting. Those in the Minimizing groups were trained to inhibit their subvocal activity in accordance with the feedback: A buzzer connected to the polygraph would sound whenever their subvocalization exceeded *E*'s criterion. Those *S*s who had been found to subvocalize during reading were given training on that task. Once *S* reached criterion, the next task was given.

For the actual task, each word triad was presented for 1.5 sec. over a TV monitor. During the 15-sec. retention interval, *S*s were instructed to work on figure analogies (similar to those on IQ tests) provided, and then to try to recall the words when "Recall" appeared on the screen. Recall was written in answer booklets, arranged so that previous answers were not visible to *S*. The recall interval was 20 sec., with a warning signal 3 sec. before the next triad appeared. Hence, the ITI was approximately 40 sec. The *S*s in the Minimizing conditions were told to keep the buzzer from sounding from item presentation until recall.

RESULTS

Due to equipment trouble and failures of *S*s to follow directions, data from 11 *S*s were discarded. The *S*s in the four cells (Similarity × Minimizing) did not differ on either pretrial performance or number of figure analogies completed. Recall was analyzed as correct/incorrect for each trial for each *S* (any error meant a trial was called incorrect), yielding a 2 × 2 × 2 × 6 × 6 matrix with one binary observation per cell. An ANOVA was done using the analysis recommended by Murdock and Ogilvie (1968). Since neither replications nor any of the interactions involving replications were significant, all other analyses were done on a collapsed matrix, yielding a 2 × 2 design (Similarity × Minimizing) with 12 *S*s per cell; trials (6) was a within-*S* factor. Mean probability of recall for the four cells was as follows: HS, yes = .67; HS, no = .61; LS, yes = .88; LS, no = .88. Similarity was highly significant, F (1, 44) = 25.74, $p < .001$, as was trials, F (5, 220) = 3.50, $p < .005$. The Similarity × Trials interaction was also significant, F (5, 220) = 2.28, $p < .05$. While the low-similarity condition showed essentially stable performance, the high-similarity group showed the zig-zag function typical of a PI-release paradigm (Wickens, 1970). In this case, however, PI buildup occurred over only a single trial (see Fig. 1). Neither Minimizing nor any of the interactions involving Minimizing approached significance. The finding that minimizing subvocal activity did not affect overall recall suggests that, if anything, the competing-response technique of minimizing "articulation" may produce interference as well as reduce subvocal rehearsal (cf. Peterson & Johnson, 1971).

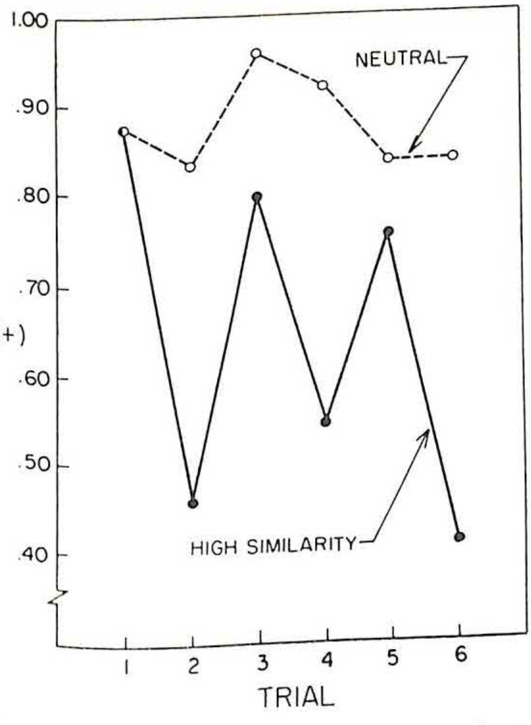


FIG. 1. Probability of correct recall, summed over Minimizing. (n at each point = 24.)

TABLE 2

PROPORTIONS OF ACOUSTIC ERRORS

Minimizing cond.	Similarity		$\bar{X}$
	High	Low	
Yes	.66 (11)	.10 (5)	.48
No	.84 (11)	.80 (5)	.83
$\bar{X}$	.75	.45	

Note.—In each case, number of *S*s on which cell mean is based is given in parentheses.

A detailed analysis of errors was also done. Since there were three words per trial, *S* could make from zero to three errors per trial. Errors were scored as omissions, intrusions, acoustic errors, transpositions, and insertions (of extralist items). While in some cases an error might be scored twice (e.g., a Trial 2 error consisting of a Trial 1 word in a high-similarity condition would be both an intrusion and an acoustic error), only one error would be scored in the analysis (in the case above, an acoustic error). It was found that on four trials, two in each of the low-similarity conditions, *S*s had skipped a response, and reference to *E* comments on the subvocalization records showed that *S*s may have failed to observe the items at presentation. These four trials were not scored as errors in the analysis below (although they were in the recall data). Since analysis of particular types of errors requires that *S* has made some error, and not all *S*s did, tests were done using an incomplete-N design for analysis of variance (Winer, 1962), with Similarity and Minimizing conditions reviewed. The frequency of acoustic errors is the prime datum, with acoustic errors accounting for 58 of the 83 errors. It might be argued that more acoustic errors would occur with material of high acoustic similarity, and also when subvocalization occurs. The tests (using square-root transformation on the raw data) support these predictions: Many more acoustic errors occurred with high-similarity items, F (1, 28) = 10.64, $p < .005$, and there also tended to be more when *S*s were not trained to minimize subvocalizing, F (1, 28) = 6.73, $p < .025$. The finding that the number of acoustic confusions relates to similarity probably represents a confound, in that with high-similarity items there are more *opportunities* for acoustic errors, relative to low-similarity items. A related question concerns the proportion of errors that are acoustic. If subvocal activity represents the source of confusions, Minimizing should reduce this proportion. To examine this, a score, P (acoustic/error), was calculated

TABLE 3

PROPORTIONS OF ACOUSTIC ERRORS, OMITTING INTRUSIONS IN THE HIGH-SIMILARITY CONDITIONS

Minimizing cond.	Similarity		$\bar{X}$
	High	Low	
Yes	.31 (8)	.10 (5)	.23
No	.71 (10)	.80 (5)	.74
$\bar{X}$	.53	.45	

Note.—Number of *S*s on which cell means are based are given in parentheses.

for each *S*, and then these scores were tested using an incomplete-*N* ANOVA; the cell means and *N*s are given in Table 2. Minimizing clearly affected the proportion of acoustic errors, $F\ (1,\ 28) = 13.10$, $p < .005$. Both Similarity and the Similarity × Minimizing interaction were weakly significant; for Similarity, $F\ (1,\ 28) = 6.20$, $p < .05$, and for the interaction, $F\ (1,\ 28) = 4.82$, which just passes the .05 level. The interaction effect should be considered suspect, given its significance level, and the effect of Similarity may represent the same confound that is present in the test of frequency of acoustic confusions. The point here is not to deny the obvious effect of Similarity on recall but to note that the intrusion errors are ambiguous: They could imply temporal confusion as easily as acoustic confusion. If one considers only spurious acoustic errors—that is, cases in which *S* makes an acoustic error which does not consist of the previous item—the pattern is clearer. Table 3 gives the same data as Table 2 except that intrusions in the high-similarity conditions have been omitted. Minimizing significantly affected the proportion of acoustic errors, $F\ (1,\ 24) = 12.87$, $p < .005$. Neither Similarity nor the Similarity × Minimizing interaction approached significance. These error analyses suggest that while the difficulty of an STM task is affected by the similarity of items, the degree of acoustic confusion is determined by the degree of subvocalization.

Beyond visual inspection of the polygraph output, no analysis of subvocalization recordings was done. While *S*s in the Minimizing conditions generally complied with feedback, activity occasionally exceeded the criterion. In all cases, the average activity of minimizing *S*s seemed to be below that of control *S*s.

DISCUSSION

These data suggest that subvocal activity mediates the production of acoustic errors in STM. The ANOVA of overall recall revealed that Similarity had a strong effect, while Minimizing did not; the error analyses showed that Minimizing had a greater effect on the nature of errors than did Similarity. With regard to the alternate hypotheses discussed earlier, the results seem consistent with the fourth hypothesis, that is, that subvocalization represents only one of several forms of storage in STM, and minimizing subvocal activity can lead to a drop in acoustic errors without affecting overall recall, due to *S* shifting to another mode of storage. Before fully accepting this, one must consider a potential problem. It is not necessarily the case that this experiment is sensitive enough to discriminate between Hypotheses *c* and *d*, that is, to decide whether or not subvocal activity mediates all forms of encoding. Presumably, such a process would occur primarily at presentation, and hence would not be greatly affected by minimizing activity during the retention interval. Neither the criterion for minimizing nor the visual measurement of activity is sensitive enough to rule out the possibility that such activity was occurring. While it is hoped that future studies may clarify this question, at present one must allow that subvocal activity may mediate all forms of encoding.

In summary, this study provides evidence for the role of subvocal activity in producing acoustic confusions in STM. In this regard, it supports the earlier suggestions of Conrad (1964) and Hintzman (1967) but goes further in specifying both the locus and general characteristics of the speech-motor mechanism. While it does not rule out auditory factors (Sperling & Speelman, 1970), it does place the burden of proof on such an interpretation to demonstrate acoustic errors in the absence of subvocal activity. At present, the important questions seem to concern alternate modes of encoding, not alternate means for producing errors.

REFERENCES

CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

GUMENIK, W. E. Effects of articulatory activity and auditory, visual, and semantic similarity on the short-term memory of visually presented paired associates. *Journal of Experimental Psychology*, 1969, **82**, 70–74.

HARDYCK, C. D., PETRINOVICH, L. F., & ELLSWORTH, D. W. Feedback on speech muscle activity during silent reading: Rapid extinction. *Science*, 1966, **154**, 1467–1468.

HINTZMAN, D. L. Articulatory coding in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 312–316.

LEVY, B. A. Role of articulation in auditory and visual short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 123–132.

LOCKE, J. L. Subvocal speech and speech. *Asha*, 1970, **12**, 8–14.

LOCKE J. L., & FEHR, F. S. Subvocal rehearsal as a form of speech. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 495–498.

MURDOCK, B. B., JR., & OGILVIE, J. C. Binomial variability in short-term memory. *Psychological Bulletin*, 1968, **70**, 256–260.

MURRAY, D. J. Vocalization-at-presentation and immediate recall, with varying recall methods. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 9–18.

MURRAY, D. J. The role of speech responses in short-term memory. *Canadian Journal of Psychology*, 1967, **21**, 263–276.

PETERSON, L. R., & JOHNSON, S. T. Some effects of minimizing articulation on short-term retention. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 346–354.

SHULMAN, H. G. Similarity effects in short-term memory. *Psyhcological Bulletin*, 1971, **75**, 399–415.

SPERLING, G., & SPEELMAN, R. G. Acoustic similarity and auditory short-term memory: Experiments and a model. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

WICKELGREN, W. A. Auditory or articulatory coding in verbal short-term memory. *Psychological Review*, 1969, **76**, 232–235.

WICKENS, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received March 20, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 170–175

EFFECTS OF INTERPOLATED TASKS ON LATENCY AND ACCURACY OF INTRAMODAL AND CROSS-MODAL SHAPE RECOGNITION BY CHILDREN[1]

SUSANNA MILLAR[2]

University of Oxford, Oxford, England

Intramodal and cross-modal shape recognition by three age groups was tested under simultaneous inputs, unfilled delay, and two types of task-filled delays. Small but consistent differences between task-filled and unfilled delays were found for intravisual but not for intrahaptic recognition at all ages. Effects on cross-modal matches were inconsistent. Preschool *S*s showed effects of task-filled delays only with visual interpolations. Visual-visual matches were more accurate and quicker than haptic-visual matches, but visual-haptic matches were neither more accurate, nor faster than haptic-haptic matches. The results are consistent with hypotheses of coding differences between modalities. It is argued that these could make cross-modal matches more liable to changes in strategy and make them less efficient.

The processes underlying matching information within and across sensory modalities are not yet well understood, and the results to date are far from consistent. Some studies, for example, suggest that cross-modal errors merely reflect errors in the respective intramodal tasks (Cashdan, 1968; Garvill & Molander, 1969—for adults; Milner & Bryant, 1970—for children; Rudel & Teuber, 1964, 1971). Others, however, show worse cross-modal than intramodal performance (Chase & Calfee, 1969—for adults; Connolly & Jones, 1970—for children; Jones & Connolly, 1970; Kress & Cross, 1969; Legge, 1965; Rubinstein & Gruenberg, 1971).

Expanding the concept and the measures of efficiency in matching may provide a way of clarifying the underlying processes. Most studies, for example, have used measures of accuracy only. Latencies might provide additional information, especially for the condition of visual-visual matching where accuracy is often at ceiling level. In addition, most studies use measures obtained under conditions of simultaneous or near-simultaneous exploration of standard and comparison. Milner and Bryant (1970), however, have suggested that cross-modal errors are greater than intramodal errors only when matching is delayed, and three studies have reported interactions between modalities and activities during the delay period (Goodnow, 1971; Jones & Connolly, 1970; Posner, 1967). These latter studies all point to the possibility of differences between modalities in retention or coding characteristics, although the nature of the difference varies from study to study. Posner, for instance, proposed a general difference between visual and kinesthetic judgments in coding properties, on the grounds that task-filled delays affect visual but not kinesthetic judgments of distance. Goodnow suggested a similar difference between vision and active touch, with the latter more likely to be disturbed by an increase in the number of comparison objects. She also proposed that some differences may stem from more effective retention or retrieval of visual standards. In contrast, Connolly and Jones (1970) have emphasized the response modality. Their model assumes that input information is translated into the response modality prior to being held in modality-specific short-term stores. If the model holds generally, effects of task-filled delays on the response modality should be found for haptic (active touch) and visual cross-modal performance.

As the nature of coding or the effect of delay emerges as a critical variable, more attention must be paid to the nature of any activity during the delay period. Brooks (1968), for example, suggested that visual recall by adults may be more disturbed by visual than by verbal distraction. In addition, it becomes more important to study matching performances by children. With children, the nature of the distractor, and, in particular, the effect of a visual distractor may be especially clear, in the sense that children appear more likely than adults to rely heavily on visual coding. Such reliance has been suggested by Bruner (1966) and is in line with Conrad's (1971) report of relatively little acoustic coding by preschoolers and the results of several studies indicating that verbal coding is not necessary for accurate cross-modal matching (Davenport & Rogers, 1970; Millar, 1971; Rudel & Teuber, 1964, 1971). Visual as well as verbal distractors should thus be used.

---

[1] This study was supported by a grant from the Social Science Research Council which is gratefully acknowledged.

[2] Requests for reprints should be sent to S. Millar, Department of Experimental Psychology, South Parks Road, Oxford OX1 3PS, England.

The present study, therefore, compares latency and accuracy of visual and haptic intramodal and cross-modal matches by children of different ages under simultaneous inputs (where accuracy is assumed to be highest), under unfilled delay, and with two types of distractors, to test three hypotheses that arise from previous findings: (*a*) Visual but not haptic matches differ under task filled compared to unfilled delays; (*b*) Task-filled delays affect the response modality in cross-modal matches; (*c*) Preschoolers' visual matches are disturbed more by visual than by verbal distractors.

METHOD AND DESIGN

A matching-to-sample task with three choice items was presented in a 3 × 4 × 4 factorial design with 3 age groups, 4 interval conditions, and 4 modality (within-*S*) conditions as variables.

Age groups.—The *S*s were 108 children from the same urban catchment area: Group 1, mean age 8 yr. 5 mo. (7 yr. 7 mo. to 9 yr. 4 mo.); Group 2, mean age 6 yr. 0 mo. (5 yr. 4 mo. to 6 yr. 8 mo.); Group 3, mean age 3 yr. 11 mo. (3 yr. 0 mo. to 4 yr. 6 mo.). An equal number of boys and girls were allocated randomly to the four interval conditions at every age level.

Interval conditions.—(*a*) Simultaneous presentation: Standard and comparison stimuli were presented at the same time on separate trays. The other three interval conditions required delayed recognition. Standards were exposed for 5 sec., followed by 9-sec. delay which differed as follows: (*b*) Unfilled delay: The *S* simply waited through the delay; (*c*) Digit repetition: Digit strings of lengths equal (and greater or less by one digit) to the *S*'s digit span, were presented aurally and repeated by the *S* during the delay. (*d*) Visual memory task: During the delay the *S* viewed highly colored, complex cartoon-type pictures. Instructions implied, but did not state, that *S*s might be tested on remembering all details. No tests were given.

Modality conditions.—Standard and choice items were presented either visually or haptically as follows: (*a*) visual-visual, (*b*) haptic-visual, (*c*) visual-haptic, (*d*) haptic-haptic. Modality conditions were run in blocks for each *S*, the order being counterbalanced across *S*s and interval conditions.

Stimuli.—Stimuli were four sets of three white Perspex nonsense shapes (2.5 × 2.5 × 1.3 cm.), described elsewhere (Millar, 1971). The order of sets was randomized across *S*s, interval, and modality conditions.

Each shape in a set served as standard twice (six responses).

Apparatus.—Stimuli, mounted on metal pins to fit into spring contacts on black standard and choice trays, were exposed in a (43 × 51 × 30.5 cm.) unit. Adjacent shuttered windows on the sloping front panel gave *S*s unrestricted views of the standard (left) and choice (right) trays, respectively, in the appropriate conditions. The shutters were linked to solenoids, automatically activated by control circuits in a separate timing and programming unit. This also controlled cuing lights for *E*. Tactual access was via two sleeved openings below the viewing windows. Standards were presented on the left. Response was by lifting one of the three shapes by its pin from the choice (right) tray. (In visual conditions, stimuli were encased in transparent Perspex to prevent accidental cues when lifting the chosen stimulus.) Response latencies were recorded on a second timer connected to the apparatus. A 6-mm. lift was sufficient to break the circuit.

Procedure.—The *S* was seated at a comfortable viewing height with hands resting beside the stimulus trays. Trials with a cross, square, and circle were run to three errorless responses prior to each modality condition. The *S* was informed of changes. Simultaneous presentations were signaled by "look and find" ("feel-look and find"/"look-feel and find"/"feel and find"). In delay conditions, standards were signaled by "look" ("feel"); and matching, by "find."

RESULTS

Accuracy

Overall results are presented in Table 1. Since accuracy varied considerably with age, correct responses were subjected to separate 4 × 4 (Interval Conditions × Modality, repeated measures) analyses of

TABLE 1

PROPORTION CORRECT CHOICES (CHANCE = .33) AND MEAN LATENCIES (LOG SEC.) FOR CORRECT CHOICES

Age group and interval cond.	Modality							
	Visual-visual		Haptic-visual		Visual-haptic		Haptic-haptic	
	Proportion correct	Mean latencies (log sec.)	Proportion correct	Mean latencies (log sec.)	Proportion correct	Mean latencies (log sec.)	Proportion correct	Mean latencies (log sec.)
Group 1:								
Simultaneous	1.00	.42	.85	.75	.64	.81	.82	.93
Delayed								
Unfilled	.92	.43	.69	.62	.80	.93	.82	.96
Filled-digit repetition	.94	.55*	.72	.60	.60	.93	.78	.95
Filled-visual memory	.97	.56*	.78	.61	.71	.96	.76	.92
Group 2:								
Simultaneous	.94	.59	.69	.87	.68	.98	.57	.95
Delayed								
Unfilled	.93	.59	.64	.69	.57	.95	.56	.99
Filled-digit repetition	.83*	.67	.49	.83*	.61	.99	.56	.93
Filled-visual memory	.93	.54	.57	.82	.54	.97	.47	.89
Group 3:								
Simultaneous	.96	.89	.53	.98	.49	1.15	.42	.80
Delayed								
Unfilled	.71	.92	.38	.81	.32	.87	.36	1.14
Filled-digit repetition	.56	.87	.33	.87	.46	.83	.43	1.02
Filled-visual memory	.49*	.80	.35	.89	.50	1.06	.38	1.20

Note.—Age range = 7 yr. and 9 mo. to 9 yr. and 4 mo., 5 yr. and 4 mo. to 6 yr. and 8 mo., 3 yr and 0 mo. to 4 yr. and 6 mo. for Groups 1, 2, and 3, respectively.

* Different from unfilled delay($p < .05$).

variance for the three groups. Results, followed by t tests of the comparisons between unfilled and task-filled delays that test the three experimental hypotheses, are set out below.

Group 1 (7 yr. 7 mo.–9 yr. 4 mo.).—Visual-visual accuracy was at ceiling level. Excluding these scores, modality, F (2, 88) = 3.65, $p < .05$, was a significant effect, due to less accurate visual-haptic than haptic-haptic matches ($p < .05$ on Newman-Keuls tests). Lower visual-haptic scores occurred mainly with simultaneous inputs and digit repetition, but neither differed significantly from unfilled delay. Haptic-visual and haptic-haptic matches did not differ. The overall picture was of lower scores on one of the cross-modal compared to either of the intramodal matches, and highly accurate visual-visual performance.

Not surprisingly in view of their high accuracy, visual-visual matches did not differ between task-filled and unfilled delays. As expected, haptic-haptic matches did not differ between conditions. However, haptic-visual matches showed no differences either. If anything, the tendency was in the opposite direction to that predicted on Connolly and Jones' (1970) model: The difference between digit repetition and unfilled delay for visual-haptic matches was just below significance level on a two-tailed t test.

Group 2 (5 yr. 4 mo.–6 yr. 8 mo.).—Modality, F (3, 132) = 39.3, $p < .001$, was the only significant effect. This meant that visual-visual matches were significantly more accurate than all others ($p < .01$ on Newman-Keuls tests). The two cross-modal matches did not differ from each other or from haptic-haptic matches.

On t tests, visual-visual matches were significantly less accurate under digit repetition ($p < .05$), while haptic-haptic matches did not differ on any comparison. Haptic-visual matches were somewhat but not significantly less accurate under digit repetition.

Group 3 (3 yr. 0 mo.–4 yr. 6 mo.).—Interval conditions, F (3, 44) = 3.79,

$p < .025$; modality, F $(3, 132) = 18.5$, $p < .001$; and an Interval Conditions × Modality interaction, F $(9, 132) = 2.83$, $p < .01$; were significant effects. Interval conditions were significant only for visual-visual matches, $p < .001$ (simple effects). Simultaneous inputs produced better matches than unfilled delay, $p < .01$, and above chance cross-modal scores, $p < .01$. However, visual-haptic matches were above chance ($p < .01$) also in at least one delay condition (visual task). Thus delay did not seem to penalize cross-modal performance specifically. Haptic-haptic and cross-modal matches did not differ from each other.

The significant modality and interaction effects meant that visual-visual matches were superior to all others, but only with simultaneous inputs, $p < .001$, and unfilled delay, $p < .001$ (simple effects), not under task-filled conditions; thus showing an effect of the latter.

On t tests, visual-visual matches were significantly less accurate only under visual tasks compared to unfilled delay ($p < .05$); the difference for digit repetition was not significant. Haptic-haptic matches did not differ; neither did haptic-visual (or visual-haptic) matches.

Latencies

Latencies (log seconds) for correct responses are shown in Table 1. Since haptic recognition involved time-consuming manual exploration, latencies for matches requiring visual recognition and latencies for matches requiring haptic recognition were subjected to separate 4 × 2 (Interval Conditions × Modality of the Standard, repeated measures) analyses of variance. (Cells with 0 correct responses were substituted by mean values, and degrees of freedom reduced accordingly; Winer, 1962.) Results, followed by t tests of comparisons testing the three experimental hypotheses, are presented below for the two older groups. For Group 3 accuracy was too low in many matches for latencies to be meaningful.

Group 1 (7 yr. 7 mo.–9 yr. 4 mo.).—Analysis of visual-visual and haptic-visual latencies showed modality of the standard, F $(1, 44) = 30.62$, $p < .001$ and an Interval Conditions × Modality of the Standard interaction, F $(3, 44) = 6.12$, $p < .01$, to be significant effects. This meant that visual-visual matches were faster than haptic-visual matches, but only with simultaneous inputs, $p < .001$, and unfilled delay, $p < .01$ (simple effects), not under either of the task-filled delays; thus showing an effect of the latter in the predicted direction.

The 4 × 2 analysis of variance of visual-haptic and haptic-visual latencies did not show effects of either interval conditions or modality of the standard.

Thus, while visual recognition was faster for visual than for haptic standards, except under task-filled delays, haptic recognition was no faster for visual than for haptic standards.

The t test comparisons confirmed that visual-visual matches were significantly slower under digit repetition, $p < .05$, and under visual tasks, $p < .05$, than under unfilled delays. As expected, haptic-haptic matches did not differ between conditions. Contrary to prediction, haptic-visual matches did not differ between task-filled and unfilled delays.

Group 2 (5 yr. 4 mo.–6 yr. 8 mo.).—Analysis of visual-visual and haptic-visual latencies showed significant effects of the modality of the standard, $F\,(1, 42) = 74.18$, $p < .001$, and of the interaction between interval conditions and modality of the standard, F $(3.42) = 3.43$, $p < .05$. The modality effect was due to faster visual-visual than haptic-visual matches. The interaction meant that interval conditions were significant only for haptic-visual matches, $p < .001$ (simple effect).

Analysis of visual-haptic and haptic-haptic latencies showed no effects of interval conditions or modality of the standard.

Visual-visual matches were somewhat, but not significantly, slower under digit repetition than under unfilled delay. Haptic-haptic latencies did not differ. Haptic-visual latencies were significantly longer under digit repetition, $p < .05$, and under visual tasks (not significant) than under unfilled delay.

Discussion and Conclusion

The pattern of intramodal and cross-modal scores was similar under simultaneous inputs and unfilled delays. Visual-haptic matches by Group 1 were somewhat less accurate and quicker, and their haptic-visual matches more accurate and slower than under unfilled delay. But differences were not significant. Unfilled delays did not penalize cross-modal matches as such. Only the youngest *S*s were significantly more accurate with simultaneous inputs, but mainly in visual-visual matches. At all ages, and in both conditions, visual-visual matches were the most accurate, and haptic-haptic did not differ from cross-modal matches. This might merely mean that cross-modal matches were as accurate as the less discriminable of the two inputs (the haptic) permitted. Similarly, the significantly shorter visual-visual than haptic-visual latencies might be due solely to better initial processing and quicker retrieval of visual than of haptic standards, as implied by Goodnow (1971). But if so, haptic recognition should also have been quicker for visual than for haptic standards. This was not the case. Visual-haptic matches were neither more accurate nor quicker than haptic-haptic. Cross-modal matches were, therefore, less efficient than could be expected from the initial processing of the standard.

To assess if coding differences between modalities might be implicated in less efficient cross-modal performance, evidence for the three experimental hypotheses must be considered. Effects of task-filled delays had been predicted for visual-visual, and none for haptic-haptic matches. Visual-visual matches by Group 1 were highly accurate but significantly slower under both task-filled delays; Group 2 were significantly less accurate and somewhat slower under digit repetition; Group 3 were significantly less accurate under visual tasks. The effects were small, but consistent, in the predicted direction, and occurred at all ages. The null hypothesis of no effects on visual-visual matches should not, therefore, be accepted. Haptic-haptic matches did not differ in either accuracy or latency at any age. Haptic matches might, of course, be disturbed by haptic interpolations. This needs investigating. The findings here are consistent with the hypothesis that visual and haptic recognition differ. On the assumption that effects of distractors are on processing or coding, this suggests coding differences.

Results did not support the second hypothesis that task-filled delays consistently affect the response modality in cross-modal matches. Only Group 2 showed the predicted effect on haptic-visual matches. The oldest *S*s were, if anything, less accurate in visual-haptic matches. Effects on cross-modal matches were thus not "fixed." This is in line with a previous result (Millar, 1972) which showed that asymmetries in cross-modal errors can change with differences in utilizing (amodal) information.

The significant decrement in preschool *S*s' visual-visual matches under visual task conditions only, is consistent with the third hypothesis. It might explain why their visual-visual matches were less accurate with delay: Visual codes may be less efficient. For older *S*s', digit repetition was equally or more effective as a distractor, suggesting that intravisual matches are not necessarily "sensory," and that not only cross-modal matches may need "associative links," as suggested by Chase and Calfee (1969).

An alternative hypothesis is that cross-modal matches are limited by differences in discriminability and coding between modalities because these require extra decisions; for instance, on which of the two inputs to rely more, and whether to ignore added information from the "better" modality. The number and direction of such decisions would depend not only on coding strategies available to *S*s at different ages, but also on coding differences between modalities which are not age dependent. Cross-modal matches would thus be more liable to strategy changes. This would account for inconsistent effects of distractors, changes in asymmetries of errors (Millar, 1972), possible criterion changes, and the frequently reported greater variability of cross-modal performance. Moreover, extra decisions take time, thus explaining the relatively longer cross-modal latencies found here, and in some studies on adults (Chase & Calfee, 1969).

The present findings on the effects of task-filled delays are suggestive rather than conclusive, but they are consistent with the hypothesis that modalities may differ in coding or processing. It is argued that this could produce less efficient cross-modal performance.

REFERENCES

Brooks, L. Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, 1968, 22, 349-368.

BRUNER, J. S. On cognitive growth. In J. S. Bruner et al. (Eds.), *Studies in cognitive growth*. New York and London: Wiley, 1966.

CASHDAN, S. Visual and haptic form discrimination under conditions of successive stimulation. *Journal of Experimental Psychology*, 1968, **76**, 215–218.

CHASE, W. G., & CALFEE, R. C. Modality and similarity effects in short-term recognition memory. *Journal of Experimental Psychology*, 1969, **81**, 510–514.

CONNOLLY, K., & JONES, B. A developmental study of afferent–reafferent integration. *British Journal of Psychology*, 1970, **61**, 259–266.

CONRAD, R. The chronology of the development of covert speech in children. *Developmental Psychology*, 1971, **5**, 398–405.

DAVENPORT, R. K., & ROGERS, C. H. Intramodal equivalence of stimuli in apes. *Science*, 1970, **168**, 279–280.

GARVILL, J., & MOLANDER, B. Intra-modal and cross-modal accuracy in a form discrimination task. *Umea University Psychological Reports*, 1969, No. 11, 1–8.

GOODNOW J. J. Eye and hand: Differential memory and its effect on matching. *Neuropsychologica*, 1971, **9**, 89–95.

JONES, B., & CONNOLLY, K. Memory effects in cross matching. *British Journal of Psychology*, 1970, **61**, 267–270.

KRESS G., & CROSS, J. Visual and tactual interaction in judgments of the vertical. *Psychonomic Science*, 1969, **14**, 165–166.

LEGGE, D. Visual and proprioceptive components of motor skill. *British Journal of Psychology*, 1965, **56**, 243–254.

MILLAR, S. Visual and haptic cue utilization by preschool children: The recognition of visual and haptic stimuli presented separately and together. *Journal of Experimental Child Psychology*, 1971, **12**, 88–94.

MILLAR, S. The development of visual and kinaesthetic judgments of distance. *British Journal of Psychology*, 1972, **63**, 271–282.

MILNER, A. D., & BRYANT, P. E. Cross-modal matching by young children. *Journal of Comparative and Physiological Psychology*, 1970, **71**, 453–458.

POSNER, M. I. Short-term memory systems in human information processing. *Acta Psychologica*, 1967, **27**, 267–284.

RUBINSTEIN, L., & GRUENBERG, E. M. Intramodal and crossmodal sensory transfer of visual and auditory temporal patterns. *Perception & Psychophysics*, 1971, **9**, 385–390.

RUDEL, R. G., & TEUBER, H.-L. Crossmodal transfer of shape discrimination by children. *Neuropsychologica*, 1964, **2**, 1–8.

RUDEL, R. G., & TEUBER, H-L. Pattern recognition within and across modalities in normal and brain-injured children. *Neuropsychologica*, 1971, **9**, 389–399.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received March 27, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 176–183

TEST OF THE TOTAL-TIME HYPOTHESIS IN FREE-RECALL LEARNING[1]

LEO POSTMAN[2] AND LINDA WARREN

University of California, Berkeley

A test of the total-time hypothesis in free-recall learning is reported. The hypothesis asserts that the amount recalled is determined by the total presentation time (TPT), regardless of list length (LL) and rate of presentation (RP). The purpose of the experiment was to verify the validity of this principle and to examine some potential limitations on its operation. With TPT held constant at 60 sec., there were three conditions of presentation: (*a*) LL 20 and 3-sec. RP; (*b*) LL 30 and 2-sec. RP; (*c*) LL 60 and 1-sec. RP. A first test of recall was given either immediately after a single presentation of the list or after an interval of 20 min. The first test was followed by two additional study-test cycles. In accord with the hypothesis, the amount recalled on the first test was invariant across conditions after both retention intervals, i.e., the operation of the principle was found to be independent of delay. After the first test, gains as a function of practice were related directly to LL. The longer the list the more new items were added on each test, although intertrial retention was essentially invariant under the different treatments. The divergence of the learning curves is attributed to the covariation of LL and the amount of semantic and associative overlap among individual items.

This study is concerned with the total-time hypothesis in free-recall learning. The hypothesis asserts that the amount recalled is determined by the total presentation time (TPT), independently of the length of the list (LL) and the rate of presentation (RP) for individual items. As long as the product LL × RP is held constant, the number of items recalled is expected to be the same. This expectation was fully supported in an experiment by Murdock (1960) in which LL and RP were varied concomitantly while TPT remained unchanged. There were four combinations of LL and RP which yielded closely similar recall scores: 20-3 (a 20-word list presented at a rate of 3 sec. per item), 30-2, 40-1.5, and 60-1. In addition to the principle of recall invariance illustrated by these findings, Murdock (1960) advanced the hypothesis that the amount of recall is a linear function of TPT. We shall not consider further the latter hypothesis, which has been called into question by the results of subsequent investigations (Roberts, 1972; Waugh, 1967). Rather attention will be directed exclusively to the principle of recall invariance proper, which predicts equal recall for different combinations of LL and RP yielding the same TPT.

Recent experimental evidence has cast doubt on the validity of the principle of recall invariance. Varying LL and RP factorially, Roberts (1972) failed to find equal recall for those conditions in his design that were equated for TPT. At a given level of TPT, performance improved as LL and RP were increased concomitantly. As Roberts points out, the method of testing used by Murdock (1960) was not, strictly speaking, free recall since *S*s were cued alphabetically. This procedural difference may be responsible for the discrepancy in results; alphabetical cues may facilitate retrieval (cf. Earhard, 1967a, 1967b) and thus obscure differences between different combinations of LL and RP. On the other hand, *S*s in Roberts' study learned a total of 480 lists (20 lists in each of 24 experimental sessions) which were all drawn from the same pool of 500 words. There were significant practice effects which interacted with both LL and RP. His conclusions concerning the principle of recall invariance may, therefore, have limited generality. Thus, a further empirical test of the

---

[1] This research was supported by Grant MH-12006 from the National Institute of Mental Health.

[2] Requests for reprints should be sent to Leo Postman, Institute of Human Learning, University of California, Berkeley, California 94720.

principle, which is free of the complications of the previous studies, will be useful.

The present experiment was designed not only to provide a further test of the principle of recall invariance but also to examine two potential limitations on its operation. One of the questions to be considered is whether the principle holds when recall is delayed rather than immediate. It is possible that on an immediate test, recall is the same for various combinations of LL and RP but that there are differences in item strength or list organization to which the test is not sensitive. Such differences should be brought out on a delayed test.

The second question is whether the principle continues to operate beyond the first study-test cycle on a list. The relevant evidence so far is confined to single-trial free recall. If the predicted invariance is obtained after a single presentation, the implication is that the level of recall reflects directly the amount of processing time. On that assumption the number of items added after each subsequent presentation of the list should remain invariant as well, provided performance is not differentially constrained by ceiling effects. The ceiling of performance will, of course, be reached sooner on short than on long lists, so that the principle of invariance must eventually break down as practice continues. Apart from such inevitable artifacts of measurement, the question remains of whether *S*'s ability to add items to those carried over from the first trial is determined entirely by the amount of processing time. The possibility of grouping temporally adjacent or semantically related items may well be systematically influenced by both LL and RP. Such systematic effects should assume increasing importance as the number of available units must be expanded and consolidated.

To permit a consideration of the questions outlined above, the experimental procedures included (*a*) concomitant variation of LL and RP, (*b*) immediate and delayed tests of recall, and (*c*) continuation of practice beyond a single study-test cycle.

METHOD

Design.—The experimental design comprised six independent groups representing the factorial combination of three conditions of list presentation and two retention intervals. The TPT was held constant at 60 sec. The conditions of presentation were: a 20-word list exposed at a rate of 3 sec. per item, a 30-word list exposed at a rate of 2 sec. per item, and a 60-word list exposed at a rate of 1 sec. per item. These treatments will be referred to as 20-3, 30-2, and 60-1, respectively. The retention interval manipulated in the experiment was that between the first presentation of the list and the first test of recall. That test was given either immediately or after 20 min. The first test was followed by two additional study-test cycles (with the test always given immediately after the presentation of the list).

Materials.—The learning materials were two sets of 60 words chosen randomly from the 1,000 most common words in the Thorndike-Lorge count. The two sets of items were used equally often. In Cond. 60-1, a list consisted of an entire set. For purposes of constructing the lists for Cond. 20-3 and 30-2, each set was divided into six groups of 10 words each. Six different lists were then formed from a given set such that each 10-word group was used equally often, namely twice each for 20-3 and three times each for 30-2.

Procedure.—On study trials the words were presented at the appropriate rate by a Carousel slide projector. On test trials *S*s were given 3 min. to write down as many words as they could remember, in any order they wished. When the first test of recall was delayed, the retention interval was filled with work on a series of mathematical problems.

Subjects.—There were 24 *S*s in each of the six groups. The *S*s were undergraduate students at the University of California who were not necessarily naive to verbal learning experiments. Assignment to conditions was in 12 blocks, with 2 *S*s per condition in each block. The *S*s were scheduled to be run in groups of 2; however, when only 1 *S* appeared, he was run by himself and a second single *S* was run thereafter. The running order within blocks was determined by means of a table of random numbers. Within a block the same lists were always used for the immediate and the delayed conditions of testing.

RESULTS

First test of recall.—Table 1 shows the mean number of items recalled by each of the experimental groups on the first test of recall. On the immediate test there were only minor variations among the three conditions of list presentation. While there was a decline on the delayed test, the scores for the three types of list remained closely comparable. Thus, the expected invariance was observed after both retention

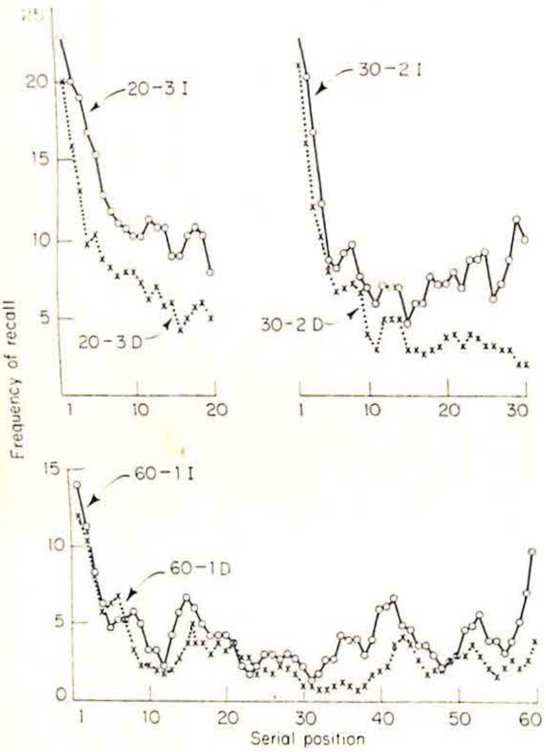


FIG. 1. Serial position curves for the three conditions of presentation on the immediate and the delayed test of recall.

intervals. This conclusion was confirmed by the results of an analysis of variance. The F ratios for list and the List × Retention Interval interaction were less than 1.00. The overall difference between the immediate and the delayed test was significant, $F(1, 138) = 60.55$, $p < .01$.

The serial position curves for the first test of recall are presented in Fig. 1. For purposes of smoothing the functions, running averages were used in plotting the curves; that is, each point, except the first and last one, represents the average of the given position, the preceding one, and the succeeding one. On the immediate test there was a strong primacy effect under all conditions of list presentation. A recency effect was present only for Cond. 30-2 and 60-1, but in both cases was less pronounced than the primacy effect.[3] When Cond. 30-2 and 60-1 are compared, the influence of recency relative to that of primacy appears to increase with LL.[4] Whatever recency effects were present in immediate recall were entirely lost on the delayed test. This finding is in agreement with the results of earlier studies showing the terminal rise in the serial position curve to be transitory (Craik, 1970; Glanzer & Cunitz, 1966; Postman & Phillips, 1965). In the case of the 20-item list, which did not yield a recency effect on the immediate test, the relative advantage of the initial over the terminal positions likewise showed some increase after a delay.

TABLE 1

SUMMARY OF MEAN RECALL SCORES

Cond.	Immediate test			Delayed test		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
20-3	10.46	15.17	17.38	7.00	14.00	16.58
30-2	11.42	16.96	19.63	6.88	14.75	19.88
60-1	11.21	21.00	24.83	7.46	19.75	24.63

The ratio $B/B + E$ was used to evaluate the change in the serial position function over time, where B refers to the score for the five positions at the beginning of the list and E, to the score for the five positions at the end of the list. In the interest of obtaining stable measures, each group was divided into six composite *S*s. The ratio for each composite S was derived from the recall scores of the individual *S*s in two successive blocks (there were two *S*s per condition in each block). The mean ratios on the immediate and the delayed test, respectively, were as follows: 20-3—.654 and .713; 30-2—.650 and .860; 60-1—.601 and .757. The higher the ratio the greater is the advantage of the initial over the terminal positions. Analysis of variance showed the increase in the ratio between the immediate and the delayed test to be significant, $F(1, 30) = 14.78$, $p < .01$. Neither the main effect of list nor the List × Re-

---

[3] A recency effect which is essentially independent of length of list has been observed with experienced *S*s (e.g., Murdock, 1962). In the present experiment there were some increases in the recency effect on the subsequent test trials, but these were neither pronounced nor regular.

[4] LL and RP were, of course, varied concomitantly. For convenience of exposition, variations in performance will be described as a function of LL.

tention Interval interaction was reliable, *F*s (2, 30) = 1.76 and 1.42.

The dominance of primacy over recency was reflected in the order of output. An index of the degree of dominance is provided by the ratio *B*/*E*, where *B* and *E* refer to the numbers of *S*s starting their recall with one of the five initial and one of the five terminal items, respectively. The ratios on the immediate test were 16/5 for 20-3, 16/3 for 30-2, and 9/5 for 60-1. On the delayed test the corresponding values were 17/1, 15/1, and 6/1. It will be noted that in the recall of the longest list the *B* and *E* segments account for a smaller number of first responses than in the case of the two shorter ones. There were, of course, more words recalled from the beginning than from the end of the list, so that more *B* than *E* items would be expected to occur in any output position. It is likely, however, that the output bias contributed to the difference in the amount recalled (cf. Deese & Kaufman, 1957). It appears, then, that during their initial exposure to the list, *S*s were disposed to rehearse the words in the order of presentation and then to start their recall with items from the beginning of the list. The only substantial departure from this pattern occurred for the 60-item list whose length would make sustained rehearsal of the items in the order of their appearance difficult; the difference in serial position effects did not, however, disturb the invariance in the total amounts recalled. As for the changes in the serial position curves after a delay, it is reasonable to suppose that the relative levels of retention of the *B* and *E* segments reflected differences in the amount of rehearsal during the input trial (Rundus, 1971; Rundus & Atkinson, 1970).

Gains as a function of practice.—The mean numbers of items recalled on the second and third trials are presented in Table 1. The scores for the three conditions did not remain invariant after the first trial, at either retention interval. Rather, the amount of gain increased directly with LL, so that there were appreciable differences in performance on both the second and third trials. Furthermore, there was rapid recovery from the losses produced by the interpolation of a retention interval between the initial presentation of the list and the first test trial. On the third trial, the scores under the immediate and the delayed treatment were closely similar.

It is apparent that the measures of recall on List 20-3 may have been constrained by the approach to the ceiling of performance, especially on the third trial. The statistical evaluation of differential gains in performance, therefore, was limited to Cond. 30-2 and 60-1, for both of which the amounts recalled remained well below the maximum. For these two conditions the main effects of both list and retention interval were significant at the .01 level in an analysis of the linear trend over trials, *F*s (1, 92) = 23.71 and 17.04. The interaction was not reliable. Both main effects were also significant ($p < .01$) in an analysis of the quadratic component of the trend over trials, *F*s (1, 92) = 8.02 and 10.94; the increases in recall showed a greater degree of negative acceleration for Cond. 60-1 than 30-2, and under the delayed than the immediate treatment.

Recall of old and new items.—For purposes of determining the source of the differences in the rate of improvement, the items in each *S*'s protocols for Trials 2 and 3 were divided into three categories: CC (correct on both the preceding and the current test trial), CN (correct on the preceding but not on the current trial), and NC (correct on the current but not on the preceding trial). These categories represent retained, forgotten, and newly recalled items, respectively. The classification was exhaustive for Trial 2; in the case of Trial 3, the assignment to categories was with reference to Trial 2; and items were pooled without regard to their status on Trial 1. The mean numbers of items per *S* falling into each of the categories are shown in Table 2. The pattern was essentially the same on both trials: The CC values were closely similar for the three conditions; however, both the CN and the NC measures rose as a function of LL. Since the increases were substantially greater for

TABLE 2

MEAN NUMBERS OF ITEMS RETAINED, FORGOTTEN, AND NEWLY RECALLED ON TRIALS 2 AND 3

Cond.	Immediate test						Delayed test					
	Trial 2			Trial 3			Trial 2			Trial 3		
	CC	CN	NC	CC	CN	NC	CC	CN	NC	CC	CN	NC
20-3	8.79	1.67	6.38	13.71	1.45	3.67	6.54	.46	7.46	12.29	1.71	4.29
30-2	8.88	2.54	8.08	14.08	2.88	6.92	5.50	1.38	9.25	12.29	2.46	7.58
60-1	7.79	3.46	11.83	14.29	5.29	10.54	6.17	1.29	13.58	13.83	5.92	10.79

Note.—CC = retained; CN = forgotten; NC = newly recalled.

NC than CN, the net improvement was directly related to length. Thus, the differential gains in performance are attributable to the rate at which new items were added.

In light of these findings, an attempt was made to determine whether *S*s tended to treat previously recalled (old) and previously not recalled (new) items as distinct categories in output.[5] To the extent that they did, the recall protocols should show categorical clustering when items are classified as CC and NC. The degree of such clustering was assessed for Trial 2, which permits a clear-cut differentiation between old and new items. The measure of clustering used was the z score difference between the observed and the expected number of runs in *S*'s recall protocol (Frankel & Cole, 1971). The mean z scores on the immediate and the delayed test, respectively, were: -1.29 and -1.57 for 20-3; -1.61 and $-.96$ for 30-2; -1.12 and -1.03 for 60-1. The differences among groups were relatively small and irregular. In each case the negative sign of the mean z score indicates the presence of categorical clustering; i.e., the observed numbers of runs fell below chance expectation. Analysis of variance shows that the overall deviation of the z scores from zero is reliable, $F\ (1, 138) = 185.83$, $p < .01$. However, none of the sources of variance associated with experimental treatments is significant. Thus, there is clear evidence for clustering based on the distinction between old and new responses, but the extent to which such grouping occurred was not influenced systematically by the conditions of presentation and testing.

The order of output of CC and NC items was considered next. The question of interest was whether there was an increasing bias toward early reproduction of new items as the list became longer. The percentage of NCs among the first half of the responses given by *S* was used to assess the degree of priority given to new items. In the absence of any bias this percentage should be equal to the percentage of NCs in the entire recall protocol. This analysis was again restricted to Trial 2. The means of the expected and the obtained percentages are presented in Table 3. The differences between expected and obtained values were in general small, although significant for the combined groups, $F\ (1, 138) = 5.56$, $p < .02$. Once more the differences among experimental conditions were neither systematic nor reliable. There was certainly no evidence that the early output of new items was favored as LL increased. It should be noted in passing that the magnitude of the priority effect was far from impressive, and the robustness of the phenomenon must remain in doubt (cf. Postman, 1972). For present purposes it is of no consequence whether the limited bias that was observed represented an artifact due to the changing order of presentation (Postman & Keppel, 1968) or a deliberate strategy on the part of *S* (Battig, Allen, & Jensen, 1965).

Intrusions.—The mean numbers of intrusions from outside the list are shown in

---

[5] We are grateful to Sheila Burns for suggesting this analysis.

TABLE 3

PRIORITY EFFECTS IN RECALL: MEAN EXPECTED AND OBTAINED PERCENTAGES OF NEW ITEMS GIVEN IN THE FIRST HALF OF RECALL ON TRIAL 2

Cond.	Immediate test			Delayed test		
	Expected	Obtained	Obtained minus expected	Expected	Obtained	Obtained minus expected
20-3	42.2	45.2	3.0	53.4	61.9	8.5
30-2	48.5	56.2	7.7	63.2	64.4	1.2
60-1	61.9	62.7	.8	69.3	70.0	.7

Table 4. The frequency of intrusions increased with LL and was higher under the delayed than the immediate conditions of testing. For each group, the number of errors declined progressively over trials. The relevant tests of significance yielded the following results: for list, $F\ (2, 138) = 10.92$, $p < .01$; for retention interval, $F\ (1, 138) = 4.30$, $p < .05$; for trials, $F\ (2, 276) = 34.54$, $p < .01$. There was also a significant Trials × Retention Interval interaction, $F\ (2, 276) = 3.50$, $p < .05$. This interaction reflects the fact that the difference between the immediate and the delayed conditions was substantially greater on the first than on the later trials. Such a localized effect of delay was to be expected since the interval was introduced between the first study trial and the first test of recall, and the influence of this delay on performance was transitory.

DISCUSSION

The principal findings were that (*a*) the measures of recall conformed to the total-time hypothesis on the first test trial, both immediately after the initial presentation of the list and after a 20-min. retention interval, and (*b*) after the first study-test cycle the amount recalled increased as a direct function of LL.

TABLE 4

MEAN NUMBERS OF EXTRALIST INTRUSIONS

Cond.	Immediate test			Delayed test		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
20-3	.88	.38	.04	1.25	.63	.54
30-2	1.04	.71	.63	2.46	1.38	.83
60-1	2.33	1.92	1.21	3.29	2.17	1.21

The results obtained on the first trial test fully confirm Murdock's (1960) original observation. Such a finding is not, as Roberts (1972) has suggested, peculiar to the procedure of alphabetical cuing used by Murdock (1960) on the test of recall. The fact remains, of course, that Roberts failed to obtain evidence for the principle of recall invariance in his own experiment. The use of a fixed pool of words in successive sessions may have influenced Roberts' results; familiarization may serve to reduce the effective differences in the amounts of material to be mastered as list length is varied. In any event, as far as single-trial free recall is concerned, the conditions limiting the operation of the principle of invariance remain to be determined. Such limitations undoubtedly exist. For example, Underwood (1970) has shown that recall is enhanced more by distributed than by massed repetitions of items within a list even though the total presentation time is held constant. In the absence of such special manipulations, however, there appear to be only minor variations in the amount recalled as long as the total study time remains the same.

The fact that the principle of invariance continued to hold after a retention interval indicates that there were no differences in the level of acquisition after a single study trial to which the immediate test was insufficiently sensitive. The retention interval was long enough to produce a significant decline in performance under all experimental treatments. Thus, variations in the degree of original learning should have been detected on the delayed test.

Why does the total-time law break down after the first study-test cycle? In the comparison of Cond. 30-2 and 60-1 an explanation in terms of differential ceiling effects can be ruled out for all practical purposes, although this factor may well have played a role in the case of Cond. 20-3. For the two longer lists the divergence of the acquisition functions must, therefore, be related to the conditions of presentation. A clue to the source of the divergence in performance is provided by the differences in the distribution of CC, NC, and CN items: The number of CCs remained invariant whereas both CNs and NCs increased directly with LL, with the latter showing a greater rise than the former. These facts suggest the following interpretation.

On the first trial, when *S* is unfamiliar with the composition and the length of the list, he is likely to rehearse words in serial order and will give relatively little attention to interitem relations. The pronounced primacy effects shown in Fig. 1 are consistent with this assumption about *S*'s rehearsal pattern on the initial study trial. Under these circumstances the number of items recalled will be determined largely by the amount of processing time. On the second and subsequent trials, *S* has become familiar with both the length of the list and the characteristics of the items included in it. He will now begin to group items on the basis of semantic and associative relations. The opportunity to form such groupings should increase with LL; the larger the number of words, the more extensive and complex the network of potential interrelations among individual words becomes. Consequently, the number of NCs increases with LL. The groupings of new items formed under conditions of rapid presentation are, however, unstable. How many of these are carried over to the next trial (CCs) once more is determined by the amount of time available for rehearsal. Hence the numbers of CCs remain invariant.

The fact that CNs as well as NCs increase as a function of list length is capable of at least two alternative interpretations. First, the more of the available study period *S* devotes to the development of new groupings, the less time he can give to the rehearsal of previously recalled items. Second, and perhaps in addition, the longer the list, the greater is the average number of items intervening between the presentation of a given old (i.e., previously recalled) word and the beginning of the test of recall. Thus, at least in the early stages of acquisition, individual items are more subject to retroactive interference when the list is long than when it is short.

The measures of clustering indicate that *S*s tended to treat previously recalled and newly acquired items as distinct categories in output. This conclusion is consistent with the finding of Robinson and Kulp (1970) that *S*s were able (*a*) to reproduce selectively the words they had recalled on a preceding test, and (*b*) to discriminate such items from unrecalled words. The occurrence of CC and NC clusters in output suggests further that during the later study trials the available time was distributed, more or less systematically, between the review of old items and the processing of new items: Words that are reproduced together are likely to have been grouped together in rehearsal (cf. Rundus, 1971). Such rehearsal and output patterns appear to have developed under all conditions of presentation; there were no reliable differences among the experimental treatments with respect to the degree of clustering. However, as suggested earlier, the ease with which NC items could be added depended on the degree of semantic and associative overlap among the words, which in turn is likely to increase with the length of the list (Deese, 1960). The fact that the frequency of outside intrusions was directly related to list length is relevant here. Such errors may be taken to represent, at least in part, the convergence of words within the list on common associates (Deese, 1959).

While the divergence in the number of NCs may reflect the advantage derived from the presence of multiple interitem associations in a long list, the invariance in the number of CCs on each successive trial points to the continuing determination of long-term storage by the amount of study time. Whether the principle of total-time invariance is found to hold will thus depend not only on the stage of practice but also on the measure of performance.

REFERENCES

Battig, W. F., Allen, M., & Jensen, A. R. Priority of free recall of newly learned items. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 175–179.

Craik, F. I. M. The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 143–148.

Deese, J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 1959, **58**, 17–22.

Deese, J. Frequency of usage and number of words in free recall: The role of association. *Psychological Reports*, 1960, **7**, 337–344.

Deese, J., & Kaufman, R. A. Serial effects in recall of unorganized and sequentially organized materials. *Journal of Experimental Psychology*, 1957, **54**, 180–187.

Earhard, M. Cued recall and free recall as a function of the number of items per cue. *Journal of*

Verbal Learning and Verbal Behavior, 1967, **6**, 257–263. (a)

EARHARD, M. The facilitation of memorization by alphabetic instructions. *Canadian Journal of Psychology*, 1967, **21**, 15–23. (b)

FRANKEL, F., & COLE, M. Measures of category clustering in free recall. *Psychological Bulletin*, 1971, **76**, 39–44.

GLANZER, M., & CUNITZ, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

MURDOCK, B. B., JR. The immediate retention of unrelated words. *Journal of Experimental Psychology*, 1960, **60**, 222–234.

MURDOCK, B. B., JR. The serial position effect of free recall. *Journal of Experimental Psychology*, 1962, **64**, 482–488.

POSTMAN, L. A pragmatic view of organization theory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

POSTMAN, L., & KEPPEL, G. Conditions determining the priority of new items in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 260–263.

POSTMAN, L., & Phillips, L. W. Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 132–138.

ROBERTS, W. A. Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology*, 1972, **92**, 365–372.

ROBINSON, J. A., & KULP, R. A. Knowledge of prior recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 84–86.

RUNDUS, D. Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 1971, **89**, 63–77.

RUNDUS, D., & ATKINSON, R. C. Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 99–105.

UNDERWOOD, B. J. A breakdown of the total-time law in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 573–580.

WAUGH, N. C. Presentation time and free recall. *Journal of Experimental Psychology*, 1967, **73**, 39–44.

(Received March 27, 1972)

ERRATUM

In the article "Temporal aspects of digit and letter inequality judgments" by John M. Parkman in the December 1971 issue, the sentence beginning on line 22, column 1, page 195, should be changed to read: "On the other hand, using the same technique, the Min accounts for an average of 53% of the variance in these two groups if the Split is partialed from the Min."

Also, in Table 3, page 198, the R^2 value for the Min under the heading "All digit pairs" should read .459.

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 184–193

PROBING FOR THE FUNCTIONAL STIMULI IN SERIAL LEARNING [1]

CARLA J. POSNANSKY [2]

University of Colorado

Probes were used to ascertain the nature of functional stimuli in serial learning (SL) revealing that (*a*) sequential, positional, compound-sequential, and other more complex cues all can function as effective stimuli in SL; and (*b*) there is considerable intraindividual variation in the use of these functional stimuli. Differences between anticipation and recall methods of SL were indicated, pointing toward more complexity within the anticipation methods. Item grouping was also shown to have differential effects upon the two methods. Reduced within-cell variances and numbers of trials to criterion were found for the recall method of SL, as was evidence for hierarchical grouping processes especially under grouped list presentations. Subsequent transfer to reversed or scrambled reorderings of the original list indicated this method to be questionable for the determination of SL processes.

Research directed at the determination of what is learned in serial learning (SL) has provided somewhat contradictory empirical results, with evidence being found for both sequential learning and positional learning as well as other more complex processes. However, the importance of each type of learning process and the degree to which each process operates independently has not been satisfactorily evaluated, and the extent to which more complex SL processes may be operating has not yet been determined (e.g., Battig, 1969b; Young, 1968). One important methodological limitation of this research is worthy of note. Transfer tasks after completed SL have provided the primary methodology for functional stimulus determination, despite frequent suggestions that a transfer task which occurs after completed SL may actually provide a measure of the utilization of serial information rather than the method by which a serial list is acquired (e.g., Battig, 1966; Postman, 1968; Voss, 1968).

The use of a transfer task after completed SL also means that the measurement of functional stimuli is occurring after various stages of learning since each item in the serial list may have been learned on a different trial. To this point, Voss (1969) was able to demonstrate that the learning of items in a serial list and the placement of those items in their proper serial positions occurred simultaneously for items at the beginning of the list. At the later positions, however, items were first learned and subsequently placed in their proper serial positions, thus demonstrating the significance of controlling degree of learning for each item. Posnansky, Battig, and Voss (1972) also reported changes in the stimulus with degree of learning by using a serial probe technique which took into account the history of the individual items, and thus differed from a similar probing technique used in serial learning by Woodward and Murdock (1968) and Woodward (1970). The present study also utilized the Posnansky et al. (1972) probe technique in an attempt to determine the functional stimulus for each item in the serial list.

The present probe technique has the further advantage of allowing the identification of functional stimuli for individual list items immediately after learning and thus permits evaluation of changes in learning process across serial positions. Such an evaluation is of particular interest given the previous contradictory results with regard to serial position (Ebenholtz, 1963; Young, 1962; Young, Patterson, & Benson, 1963). Furthermore, the probe technique permits analysis of an individual's performance which, in turn, permits the evaluation of the consistency of process within individuals as well as an evaluation of a given *S*'s functional stimulus usage in relationship to rate of SL.

---

[1] This research was supported by a National Science Foundation predoctoral fellowship to the author and by National Science Foundation Grant GH-25433 to William F. Battig. The present article is based on a thesis submitted by the author in partial fulfillment of the requirements for a master's degree and is publication number 26 of the Institute for the Study of Intellectual Behavior at the University of Colorado. The author acknowledges the assistance of the chairman of her thesis committee, William F. Battig, as well as that of Bruce Ekstrand, Verne Keenan, James Voss, and James Pellegrino.

[2] Requests of reprints should be sent to Carla J. Posnansky, Department of Psychology, University of Colorado, 0034 Muenzinger Building, Boulder, Colorado 80302.

With the use of the probe technique, a primary purpose of the present study was the evaluation of possible differences in process between the recall and anticipation methods of SL. Previous research attempts have been almost exclusively confined to the anticipation method, but Battig (1969a) reported differences in sensitivity to variable manipulations between the two methods. While these findings demonstrate that the serial recall method has greater sensitivity to these parameter manipulations, the involvement of different learning processes is not necessarily implied. Young (1968), however, pointed out that "a more critical comparison of the serial anticipation and serial recall procedures is warranted [pp. 127–128]." A major purpose of the present study was to determine if such a difference in process does exist.

The present study also included the manipulation of a within-list variable—grouping of items into three-item subgroups during learning, suggested to be a potentially important determinant of SL process in an unpublished study by Skiba and Battig (1970). Thus, a further purpose of the present investigation was the evaluation of the differential effects of such a manipulation upon SL method, as well as a direct evaluation of the variable's effect upon functional stimuli.

Finally, the present study attempted to replicate the recall and anticipation methodology differences in sensitivity to variable manipulations found by Battig and Lawrence (1967) by the inclusion of both reversed- and scrambled-order SL transfer tasks. This transfer manipulation also permits resolution of the conflict produced by the Battig–Lawrence (1967) failure to replicate the Young et al. (1963) finding of facilitation for middle items of the reverse SL task. Finally, the inclusion of such a transfer task also provided an opportunity for the evaluation of changes in the functional stimuli between original and transfer tasks, thus permitting the possible determination of process differences between task acquisition and subsequent utilization.

METHOD

Material and Apparatus

All lists consisted of 15 four-letter words of high Thorndike-Lorge frequency, including all but 1 (EAST) of the 16 maximally unrelated words representing minimal associative communality as described by Battig and Miller (1965). Four serial list orders were constructed, the first with items arranged randomly. The second list order was the reverse of the first, and the third list order was constructed so as to maintain a scrambled relationship with the first. In this scrambled list, no adjacencies or serial positions of the first list were retained. The fourth list order was the reverse of the third. Four asterisks preceded the beginning of the list, and, additionally, four asterisks succeeded the last word in the list.

Lists were presented by a Lafayette memory drum, while the probes were presented on 3 × 5 white index cards.

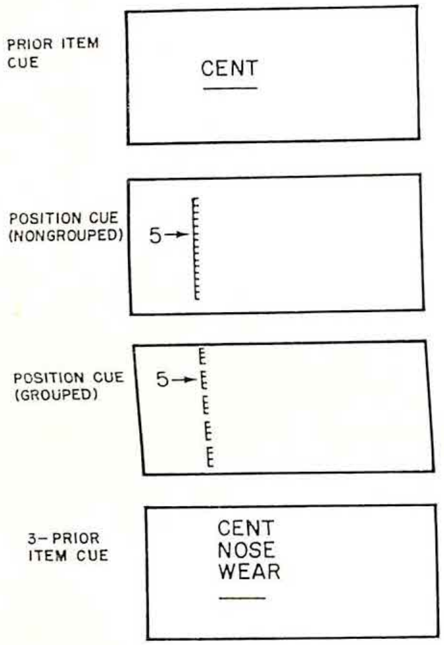
The 5-sec. probe intervals were timed by the use of a variable timer operated by foot lever.

Design and Procedure

Subjects were randomly assigned, on the basis of order of appearance in the laboratory, to one of 24 experimental cells, with the restriction that each cell be half filled before the remaining half of any other cell was filled. There were 12 experimental groups and 12 control groups, with three variables, in addition to list order, being manipulated between groups. List orders were counterbalanced so that within each cell, each of the four list orders was used equally often as an original list order and a transfer-task list order.

Variable Manipulations

Method.—There were three methods tested in the present study. The standard serial anticipation method required that *S*s study the list the first time through, and thereafter attempt to anticipate each item during the presentation of the preceding word. A 2-sec. presentation rate and a 6-sec. intertrial interval were used. Two variations of a serial recall



PROBE SEQUENCES

PI-3PI-POS GROUPED
PI-3PI-POS NONGRP
PI-POS NONGRP-3PI
PI-POS GROUPED-3PI
POS GROUPED-PI-3PI
POS NONGRP-PI-3PI

FIG. 1. Examples of probe types and probe sequences.

method both required *S*s to view the items on the memory-drum tape for 2 sec. each. Then, after the entire list had been presented, each *S* was given 1 min. to orally recall the words in the list in the order in which they had appeared. Instructions to *S*s indicated that if, at any time, *S* was aware that he was skipping a position in the list, he should indicate the nonremembered item with the word "skip," thus making it possible for words to be placed along positions without following the correct prior items. Intertrial interval was again 6 sec. A response was scored as "learned" when it was given by *S* in the correct serial position *or* when it was given by *S* following the correct preceding item under the "recall or" method, while under the "recall and" method items were counted as correct only when they were given correctly in the proper serial position *and* preceded by the correct prior item.

Grouping mode.—Lists were presented either in a continuous manner at the 2-sec. rate (nongrouped) or with a 2-sec. blank interval following every third word (grouped).

Transfer task.—All *S*s received a transfer task, after List 1 learning had been carried to a one-errorless trial criterion, which maintained either a scrambled or reversed relationship to the first list learned. Additionally, the transfer task was carried to the same criterion as original learning, and, furthermore, the probing technique to be described below was applied to the transfer task in the same manner as it was applied to original learning. Each of the four list orders was used equally often both for the reversed and the scrambled transfer conditions.

Serial Probing

The probing task represented the only within-*S*s manipulation in the present study. The probing technique is similar to that developed by Murdock and associates (e.g., Woodward & Murdock, 1968) for probing short-term memory. The technique was applied as described in Posnansky et al. (1972) to all but the first three items in the serial list. Thus, immediately after an item was scored as correct for the first time, either a diagrammatic representation of the item's serial position or its immediately preceding item was presented to *S* on an index card. Examples of the probe types are seen in Fig. 1. After card presentation, *S* was given 5 sec. in which to respond with the correct item. If *S* could not do so, an additional probe card was presented on which the alternative cue or a new cue consisting of the item's three preceding items was added to the first cue. After another 5 sec., if *S* still did not respond correctly, the third cue was added. The three-prior-item cue necessarily followed the prior-item cue in every sequence since a correct response to the three-prior-item cue without prior failure on the prior-item cue would provide no information as to which cue was effective.

As seen in Fig. 1, there were two types of diagrammatic representations of serial position so that the position cue represented the items in the list in either a divided (grouped) manner or in the continuous manner described in Posnansky et al. (1972). Each type of position cue was used equally often for each *S*, and probe sequences were counterbalanced within each experimental cell, so that the type of probe appearing first in each sequence (prior item or serial position) appeared equally often for each item across *S*s within a cell. Also, each first probe type appeared equally often with a given *S*'s 12 probed items so that all *S*s received the prior-item cue first for half of the items in the list and the position cue first for the other half of the items. These probe sequences are also shown in Fig. 1. Since each item was probed immediately following the trial on which it was given correctly for the first time, 12 probe responses were recorded for each *S* and the number of probes on any given trial matched the number of items given correctly for the first time on that trial. The sequence of probe types and the first probe responded to correctly were recorded for all probed items.

In order to permit an evaluation of the effectiveness of the probe types, a control procedure was also used in which probes were given which were noncontingent upon *S*'s trial-by-trial performance.

Items to be probed after a given trial were previously and randomly selected so that it was possible to study the effectiveness of probes for items not yet "learned." To maintain equivalence with the contingent probe conditions, the number of items probed noncontingently on each trial was equal to the number first recalled correctly on that trial.

Instructions to *S*s regarding the occurrence of probes were given with the SL instructions prior to SL. These instructions explained the use of the probe cards, the time intervals involved, and the oral responding required. No information was given, however, concerning the contingencies of probing, and the three probe types were not specifically described.

Subjects

The 112 *S*s were students at the University of Colorado who participated in order to fulfill a course requirement. The number of *S*s in each noncontingent cell was equal, while in contingent cells, the number of *S*s learning under the anticipation method was double the number of *S*s under each recall method. Because of this cell assignment, an equal-*n* comparison of the two recall and the anticipation methods within the noncontingent cells was possible. Thus, cell assignment was designed to facilitate the analysis which was of fundamental interest within each contingency condition.

RESULTS

Overall performance differences were evaluated by between-groups analyses of variance, separately for original and transfer SL tasks, using both trials and total errors to the one-errorless trial criterion as measures. Since the results were closely comparable for both measures only the trials-to-criterion data and analyses will be reported here. Additionally, data were pooled over the four list orders since an analysis of variance indicated no significant differences in learning difficulty between lists, *F*s $(3, 94) < 1.0$.

In order to evaluate probe effectiveness, three different types of analyses of variance were performed. First, an overall analysis was done separately for each of the three probe types wherein each score consisted of the number of times a given probe was responded to correctly over a pair of serial positions. It was deemed necessary to pool data across two serial positions in order to gain sufficient data for a meaningful analysis. Second, a "proportional analysis" was done to provide a fairer evaluation of the effects of the between-groups variables. Each score in the proportional analysis consisted of the number of times a given probe type was used effectively by each *S* divided by the number of occasions on which it could have been used. In this analysis, data were pooled across all serial positions. Such an analysis has the additional advantage of providing directly comparable measures representing each of the three probe types. Finally, a "stage analysis" was based on separate scores for the first, second, and third probes, i.e., each *S*'s score consisted of the number of times a given type of probe was used effectively divided by the number of opportunities for its usage within a particular stage of probing. This stage analysis was done to ascertain differential effects of the between-group variables upon probe usage after varying numbers of errors to previous probes.

The results will be organized into sections describing each of the individual variable effects both on list acquisition and probe usage. Then serial position effects will be considered as will the aspects of the present study which deal with individual differences.

Method-of-Learning Effects

The differences in number of trials to criterion between the two methods of SL are seen in Table 1 which indicates that recall *S*s needed slightly over half as many trials to reach criterion as did anticipation *S*s. The differences in variances are also worthy of note. Within the nongrouped presentation conditions, the within-cells variance for anticipation cells was threefold that of recall cells, but these variance differences disappeared within the grouped presentation mode.

Overall probe analyses indicated that the method variable had a significant effect only upon usage of the three-prior-item cue for original list probing. The *S*s learning under the anticipation method used this cue effectively for 12.5% of the items probed while recall *S*s did so for only 7.8% of the items, $F\ (1,\ 60) = 4.24$, $p < .05$. Additionally, recall *S*s used the prior-item

TABLE 1

MEANS AND VARIANCES OF ORIGINAL AND TRANSFER TASKS—TRIALS TO CRITERION

Measure	Contingent		Noncontingent	
	Grouped	Non-grouped	Grouped	Non-grouped
Original list				
Anticipation $\bar{X}$	12.63	15.38	13.00	15.25
Method variance	15.32	33.45	14.00	22.21
Recall	7.63	7.50	7.06	9.06
Method	9.45	9.20	14.26	16.86
Transfer list				
Anticipation $\bar{X}$	7.38	7.50	7.25	10.75
Method variance	14.65	30.30	9.64	37.36
Recall	4.08	5.25	4.25	6.00
Method	11.27	9.67	10.90	11.00

cue 57.3% of the times that it appeared while anticipation *S*s used the cue only on 46.7% of such occasions on the average, with F (1, 60) = 4.29, $p < .05$, in the proportional probe analysis.

There were more cases in which all three cues were ineffective in the anticipation method (25.5%) than in the recall method (20.6%), but this difference did not reach statistical significance.

Grouping Mode

As seen in Table 1, *S*s receiving original list presentation under the grouped mode needed significantly fewer mean trials (9.66) to reach criterion than did *S*s receiving continuous list presentations (11.30), F (1, 97) = 4.55, $p < .05$. This was also true in the transfer task wherein grouped *S*s needed 5.38 mean trials to reach criterion while nongrouped *S*s needed 7.38 trials. Although the Grouping Mode × Method of Learning interaction was not significant, Table 1 suggests the effect of grouping to be larger under the anticipation method. There were no main effects of the grouping mode manipulation upon probe usage.

Transfer Task Type

The *S*s transferring to a reversed serial list needed significantly fewer mean trials (4.74) to reach criterion than did *S*s transferring to a scrambled list (8.39 trials), F (1, 96) = 36.16, $p < .001$. This effect did not differ across grouping modes or methods of learning. However, it is also of interest to note that even the scrambled transfer task was learned with fewer mean trials to criterion (8.39) than were needed across all conditions of original list learning (10.48), thus indicating that the net transfer effects were not negative. All transfer means are given in Table 1.

There were no significant effects (Fs < 1.0) of transfer task type upon probe usage.

Contingency–Noncontingency Effects

Separate analyses of variance were performed for the contingent and noncontingent groups. Error variances for these two conditions were closely comparable. The method-of-learning effect was somewhat larger for the contingent group, F (1, 40) = 23.11 for noncontingent and F (1, 56) = 39.80 for contingent, both ps $< .001$. Also, the effect of the transfer condition was much greater in the contingent group, F (1, 56) = 32.17, than in the noncontingent group, F (1, 40) = 6.54, although the direction of the difference was consistent across contingency conditions.

With regard to probe usage, 77.0% of the responses by anticipation *S*s to probe sequences of "unlearned" items were never correct within the grouped list presentations, while 69.0% were never correct under the continuous presentation conditions. Under the recall procedure, 85.0% of the unlearned items probed were never responded to correctly under grouped presentations, while 74.0% of the unlearned items under continuous conditions were never probed effectively. Of course, anticipation *S*s had an additional opportunity to view an item between the incorrect anticipation of that item and the probing of that item which was designated as "unlearned." Therefore, more instances in which probes were effective for "unlearned" items were expected under anticipation conditions.

Interactions of the Between-Groups Variables

The interaction of the grouping mode and method-of-learning variables was found significant in analyses of variance for both the prior-item and three-prior-item cues. This suggests that within the recall method, the three-prior-item cue was more effective in the grouped conditons, while the opposite was true under the anticipation method. This effect was significant in the analysis of Stage 3 probing for the three-prior-item cue, $F\ (1,\ 60) = 6.77$, $p < .025$, as well as in the Stage 2 probing analysis of the prior-item cue, $F(\ 1,\ 60) = 7.10$, $p < .01$.

Probe Results

The extent to which the three probe types were effectively used differed markedly ($p < .001$), as shown in Table 2. Within original list learning, the prior-item cue was used effectively for significantly ($p < .001$) more of the items probed than was the serial position probe which was the next most effective cue in terms of frequency of usage. Additionally, the position cue was used significantly more often than the three-prior-item cue, ($p < .001$), although frequency of usage of the position cue did not differ significantly from the frequency of cases in which all three cues were ineffective. That the latter reflected inadequate learning was shown by an analysis showing significantly more errors made on subsequent trials after the probes on items for which all probes were ineffective than on items yielding correct responses to any of the probes, $F\ (1,\ 54) = 46.01$, $p < .001$.

The frequency of usage of cue types within the transfer task probing is also seen in Table 2. Here usage of the position cue was significantly ($p < .001$) greater than the frequency of cases in which none of the three cues were effective, although other probe usage frequencies were comparable to those of original list probing.

TABLE 2

PROPORTION OF PROBE USAGE

Functional stimulus	Original learning	Transfer learning
Prior item	41.0%	45.8%
Position	25.7	27.4
Three prior item	10.3	11.1
No cue effective	23.0	15.7

Serial Position Effects

An examination of the serial position curves revealed the characteristic increased bowing of the curve under the anticipation method as well as a flattening of the curves within both types of transfer tasks, indicating that the degree of positive transfer was not greater for middle serial positions within the reversed transfer condition. Additionally, an examination and analysis of serial position subgroups composed of three items each revealed that *S*s learning under the grouped presentation mode had a greater tendency to learn the last item in a three-item subgroup first than did nongrouped *S*s. Differences between first and third items within three-item subgroups were greatest at the extremes of the serial list, indicating that, in general, the greatest tendency for third items within subgroups to be learned last occurred at the beginning of the list, while the opposite was true for the last item in the list.

With regard to serial position effects upon probe effectiveness, it should be reiterated that the first three items in the list were not probed, so that references to the beginning of the serial list with respect to probe usage actually refer to the first probed positions in the list. No effects of serial position were found for the prior-item cue as seen in Fig. 1. However, the effective use of the serial position probe did vary with serial position, $F\ (5,\ 300) = 2.60$, $p < .025$, thus indicating that the cue was most effective at the ends of the list although usage did not vary across methods of learning or across grouping modes. Three-prior-item usage also did not vary significantly as a function of serial position. However, a significant Serial Position × Grouping Mode interaction did indicate that the three-prior-item cue was most effective for middle items of the list for *S*s learning under continuous list presentations, while for *S*s receiving grouped item

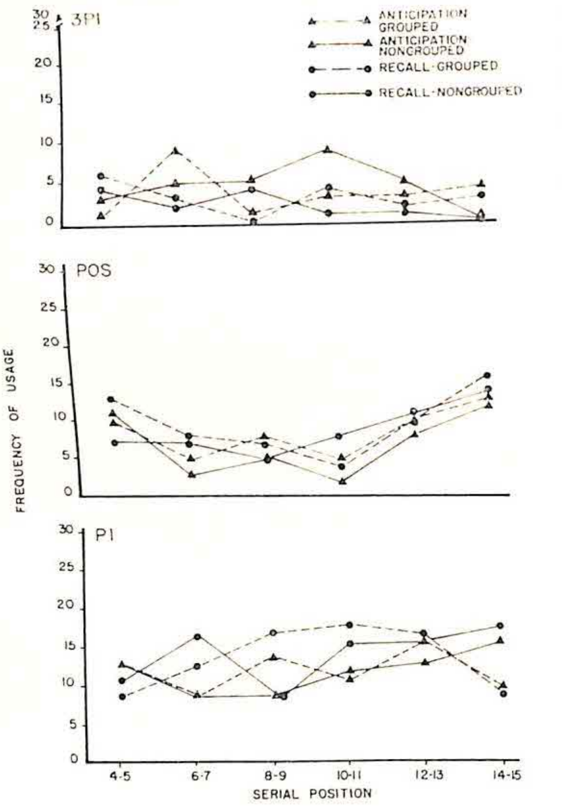


FIG. 2. Probe usage during original list probing.

presentation, the three-prior-item probe was most effective at the list's extremes, F (5, 300) = 8.26, $p < .001$. Furthermore, within the three-prior-item overall analysis, the interaction of serial position and method of learning was significant, F (5, 300) = 2.48, $p < .025$, suggesting that the three-prior-item was least effective under anticipation for the very center and end serial positions while it was most effective at the intermediate positions. In contrast, under the recall procedure, the three-prior-item cue usage was greatest at the beginning serial positions, with usage being attenuated across subsequent positions.

The occurrence of cases in which all three probe types were ineffective was most pronounced for Positions 6–11, F (5, 300) = 5.58 $p < .001$.

Probe usage for the transfer task also revealed that the position cue was especially effective at the list's latter serial positions, F (5, 315) = 11.27, $p < .001$, while the three-prior-item cue was especially ineffective for the last two positions of the list and for Positions 6 and 7, F (5, 315) = 3.85, $p < .001$.

Other Results

Several correlational analyses were performed in order to ascertain the relationship between *S*s' probe performance and ease of list acquisition. Results indicated that the use of simple cues (prior item or position) was negatively correlated with errors to criterion ($r_{\text{prior item}} = -.384$, $p < .01$, and $r_{\text{position}} = -.234$, $p < .10$), while the use of the three-prior-item cue was slightly but positively correlated ($r = .22$, $p < .10$).

However, an analysis of variance performed on the number of items per *S* wherein probing reached a three-prior-item stage indicated that such a stage was not reached more often under one learning method or under a particular grouping mode (both Fs < 1.0), so that differences for these two variables cannot be attributed to a subject selection bias.

Examination of individual *S* protocols also revealed that even an *S*'s predominantly used cue was effective for only 5.40 (45%) of the 12 probed items on the average, and this consistency did not vary as a function of serial position.

Furthermore, probe usage was insignificantly affected by the number of previously occurring probes, thus providing further evidence that the probing technique provides a relatively uncontaminated insight into SL processes.

Finally, with regard to the noncontingently probed items, the difference between recall methods is of interest in order to further validate the probing technique. Within the "recall or" method, wherein items were considered "learned" when they were given correctly in the correct serial position *or* following the correct prior item, 93% of the probe sequences for "unlearned" items resulted in no correct responding for continuous list presentation while 91% resulted in no correct responding for grouped list presentations. However, within the "recall and" groups, wherein items had to be given correctly in the correct serial positions *and* following the correct prior item before they were scored as correct, only 79% of the probe sequences for "unlearned" items were not responded to correctly under grouped list presentation, and only 55% of the probe sequences resulted in no correct responding under the continuous list presentations, thus providing a substantial difference between the two methods.

DISCUSSION

The results of the present study clearly point toward the existence of more than one SL process by showing at least some usage of all three types of functional stimuli studied. No single SL process was identified which operated consistently across *S*s or even across different items within the same *S*. The results further indicated that the use of the prior-item cue was most prevalent, followed by the position cue and then the three-prior-item cue, in order of degree of effectiveness. Additionally, the possible existence of still other, as yet unidentified, SL processes was also indicated by a substantial number of cases (23%) in which the three stimuli tested were all ineffective (although some indeterminate proportion of these may have represented guessed rather than actually learned items). Such cases were clearly greatest at the center positions of the list, supporting a previous suggestion that more complex processes operate primarily in the middle of the serial list (Battig, Brown, & Schild, 1964).

Additional evidence for the existence of several SL processes was found in the examination of individual *S*s' performance. Such an examination indicated that the degree to which each cue was effective did vary within as well as between *S*s. The average number of times that any given *S* utilized his favored functional stimulus was found to be less than half the number of probed items in the list, clearly indicating the operation of more than one SL process within a given *S*. Importantly, this finding further demonstrates that the existence of more than one SL process is not due to consistent between-*S* differences in SL.

The finding of differences in functional stimulus usage between methods of learning suggests a difference of process between methods, perhaps pointing toward increased complexity in the anticipation method wherein *S*s must use the same 2-sec. period to view an item, check their last response, and anticipate a succeeding item. It may indeed be the case that when recall *S*s are presented with the prior-item cue, they are able to use it effectively because it is similar to the recall cue that *S* himself produces while recalling the items in the serial list. On the other hand, anticipation *S*s may find a greater discrepancy between the prior-item cue presented during probing and the conditions of the anticipation method, wherein the prior item and additional temporal and sequential cues are present at the time of anticipation.

The interactions of grouping mode and method of learning which were reported further suggest that grouped list presentation adds to the complexity of the processes occurring in the recall method, making the more complex cue more potent. This result is consistent with the interpretations of the grouping mode effect made by Skiba and Battig (1970) and Keenan (1970) which suggest that grouping effects are due to *S* strategy changes or "hierarchical grouping" rather than an emphasis, for example, of serial position.

Additionally, these Grouping Mode × Method of Learning interactions suggest that within the anticipation method the blank spaces (presented during both study and test) themselves become functional stimuli for the first items in each group of three. This would diminish the three-prior-item cue effectiveness in those cases in which the blank was not presented on the probe cards.

The present results also indicated that the position probe was most effective at the list's extremes as found by Posnansky et al. (1972). Furthermore, position cue usage did not vary across grouping modes, thereby indicating that the effect of grouping (i.e., the reduction in trials to criterion under grouped conditions) was not in the emphasis of item position.

Thus, with respect to original list learning, the probe technique has demonstrated a difference of SL processes. As to probing in the transfer task, its differences from the original task in probing results become important with regard to appropriateness of transfer tasks for the evaluation of SL processes. Thus, the lack of the method-of-learning differences and the Grouping Mode × Method of Learning interactions which were found in original list probing, in addition to fewer (15.8%) cases in which all three probes were ineffective, suggest that serial information gained in original list learning is being utilized in the transfer task,

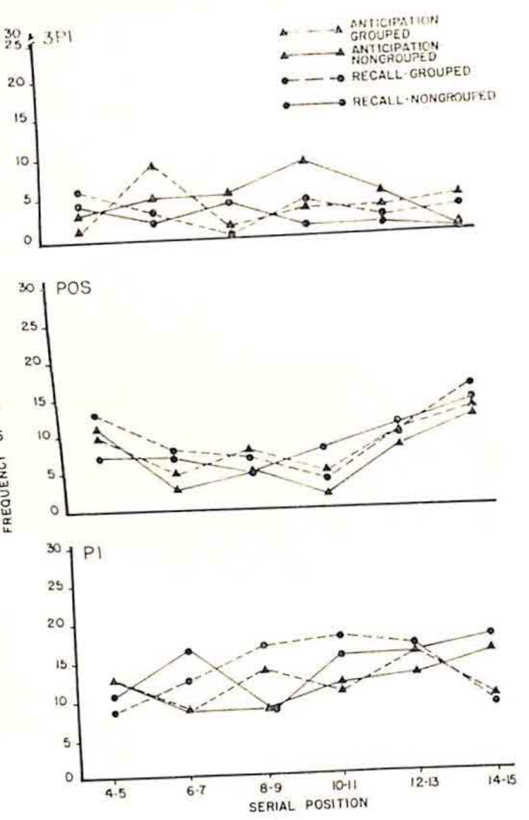


FIG. 2. Probe usage during original list probing.

presentation, the three-prior-item probe was most effective at the list's extremes, F (5, 300) = 8.26, $p < .001$. Furthermore, within the three-prior-item overall analysis, the interaction of serial position and method of learning was significant, F (5, 300) = 2.48, $p < .025$, suggesting that the three-prior-item was least effective under anticipation for the very center and end serial positions while it was most effective at the intermediate positions. In contrast, under the recall procedure, the three-prior-item cue usage was greatest at the beginning serial positions, with usage being attenuated across subsequent positions.

The occurrence of cases in which all three probe types were ineffective was most pronounced for Positions 6–11, F (5, 300) = 5.58 $p < .001$.

Probe usage for the transfer task also revealed that the position cue was especially effective at the list's latter serial positions, F (5, 315) = 11.27, $p < .001$, while the three-prior-item cue was especially ineffective for the last two positions of the list and for Positions 6 and 7, F (5, 315) = 3.85, $p < .001$.

Other Results

Several correlational analyses were performed in order to ascertain the relationship between *Ss'* probe performance and ease of list acquisition. Results indicated that the use of simple cues (prior item or position) was negatively correlated with errors to criterion ($r_{\text{prior item}} = -.384$, $p < .01$, and $r_{\text{position}} = -.234$, $p < .10$), while the use of the three-prior-item cue was slightly but positively correlated ($r = .22$, $p < .10$).

However, an analysis of variance performed on the number of items per *S* wherein probing reached a three-prior-item stage indicated that such a stage was not reached more often under one learning method or under a particular grouping mode (both Fs < 1.0), so that differences for these two variables cannot be attributed to a subject selection bias.

Examination of individual *S* protocols also revealed that even an *S*'s predominantly used cue was effective for only 5.40 (45%) of the 12 probed items on the average, and this consistency did not vary as a function of serial position.

Furthermore, probe usage was insignificantly affected by the number of previously occurring probes, thus providing further evidence that the probing technique provides a relatively uncontaminated insight into SL processes.

Finally, with regard to the noncontingently probed items, the difference between recall methods is of interest in order to further validate the probing technique. Within the "recall or" method, wherein items were considered "learned" when they were given correctly in the correct serial position *or* following the correct prior item, 93% of the probe sequences for "unlearned" items resulted in no correct responding for continuous list presentation while 91% resulted in no correct responding for grouped list presentations. However, within the "recall and" groups, wherein items had to be given correctly in the correct serial positions *and* following the correct prior item before they were scored as correct, only 79% of the probe sequences for "unlearned" items were not responded to correctly under grouped list presentation, and only 55% of the probe sequences resulted in no correct responding under the continuous list presentations, thus providing a substantial difference between the two methods.

DISCUSSION

The results of the present study clearly point toward the existence of more than one SL process by showing at least some usage of all three types of functional stimuli studied. No single SL process was identified which operated consistently across *S*s or even across different items within the same *S*. The results further indicated that the use of the prior-item cue was most prevalent, followed by the position cue and then the three-prior-item cue, in order of degree of effectiveness. Additionally, the possible existence of still other, as yet unidentified, SL processes was also indicated by a substantial number of cases (23%) in which the three stimuli tested were all ineffective (although some indeterminate proportion of these may have represented guessed rather than actually learned items). Such cases were clearly greatest at the center positions of the list, supporting a previous suggestion that more complex processes operate primarily in the middle of the serial list (Battig, Brown, & Schild, 1964).

Additional evidence for the existence of several SL processes was found in the examination of individual *S*s' performance. Such an examination indicated that the degree to which each cue was effective did vary within as well as between *S*s. The average number of times that any given *S* utilized his favored functional stimulus was found to be less than half the number of probed items in the list, clearly indicating the operation of more than one SL process within a given *S*. Importantly, this finding further demonstrates that the existence of more than one SL process is not due to consistent between-*S* differences in SL.

The finding of differences in functional stimulus usage between methods of learning suggests a difference of process between methods, perhaps pointing toward increased complexity in the anticipation method wherein *S*s must use the same 2-sec. period to view an item, check their last response, and anticipate a succeeding item. It may indeed be the case that when recall *S*s are presented with the prior-item cue, they are able to use it effectively because it is similar to the recall cue that *S* himself produces while recalling the items in the serial list. On the other hand, anticipation *S*s may find a greater discrepancy between the prior-item cue presented during probing and the conditions of the anticipation method, wherein the prior item and additional temporal and sequential cues are present at the time of anticipation.

The interactions of grouping mode and method of learning which were reported further suggest that grouped list presentation adds to the complexity of the processes occurring in the recall method, making the more complex cue more potent. This result is consistent with the interpretations of the grouping mode effect made by Skiba and Battig (1970) and Keenan (1970) which suggest that grouping effects are due to *S* strategy changes or "hierarchical grouping" rather than an emphasis, for example, of serial position.

Additionally, these Grouping Mode × Method of Learning interactions suggest that within the anticipation method the blank spaces (presented during both study and test) themselves become functional stimuli for the first items in each group of three. This would diminish the three-prior-item cue effectiveness in those cases in which the blank was not presented on the probe cards.

The present results also indicated that the position probe was most effective at the list's extremes as found by Posnansky et al. (1972). Furthermore, position cue usage did not vary across grouping modes, thereby indicating that the effect of grouping (i.e., the reduction in trials to criterion under grouped conditions) was not in the emphasis of item position.

Thus, with respect to original list learning, the probe technique has demonstrated a difference of SL processes. As to probing in the transfer task, its differences from the original task in probing results become important with regard to appropriateness of transfer tasks for the evaluation of SL processes. Thus, the lack of the method-of-learning differences and the Grouping Mode × Method of Learning interactions which were found in original list probing, in addition to fewer (15.8%) cases in which all three probes were ineffective, suggest that serial information gained in original list learning is being utilized in the transfer task, thus providing *S*s with additional information with which to respond to the probes. This essentially corroborates the position that a transfer task may actually be measuring the utilization of serial information rather than its process of acquisition (e.g., Postman, 1968).

Furthermore, determining the effectiveness of the probing technique for tapping into SL processes raises questions concerning the analysis of probe results for probes not occurring first in the probe sequence. Thus, *S*s who have difficulty with SL also may have trouble responding to probes, reflecting some type of generalized difficulty with the kinds of tasks involved. If so, an *S* selection bias might be operating upon the stage analyses of probe effectiveness, as well as upon the evaluation of the three-prior-item cue. To deal with this question, number of errors made during original learning were correlated with overall number of probe errors made, yielding an inconclusive $r = .429$. The number of times each *S* reached a three-prior-item stage was then subjected to analysis of variance, which indicated that the three-prior-item stage was not reached more often under one method of learning or under one grouping mode. Thus conclusions concerning the effects of the between-groups variable manipulations appear uncontaminated by any type of *S* bias.

Two additional correlations, of errors to criterion both with failures to respond correctly to any probe in a sequence and errors made in Stage 1 of probing, were closely comparable in magnitude. This indicates that the original correlation is not localized at a particular stage of probing, but is rather diffuse.

The preceding discussion does, however, suggest that "slow" serial learners do indeed take longer to learn a serial list because they utilize complex functional stimuli during acquisition. Some support for this notion comes indirectly from the significant negative correlations found between the use of the simplest cues and numbers of errors to criterion, in conjunction with the slightly positive correlation found between the use of the three-prior-item cue and errors to criterion.

In view of the present probing results, the inconsistent results from previous research on "what is learned in serial learning" are hardly surprising. Since individual *S*s use more than one (and often all) of the three types of functional stimuli when learning a serial list, previous research attempts to determine the functional stimulus could hardly have produced anything other than contradictory findings. Also, any conclusions about the functional stimulus emerging from this previous research typically represent the lack of positive evidence for some such stimulus where it should have been operative (Battig, 1969b). The present study clearly demonstrates the inappropriateness of this type of conclusion.

The present results also replicated those of Battig and Lawrence (1967) with respect to the fewer number of trials to criterion and the reduced variances occurring within the recall method. Quicker learning by recall *S*s can be attributed to the lack of confounding of study time and recall time under the recall method, and the additional increase in the total time required for each recall trial. However, Battig and Lawrence also found greater differences between transfer-task types under a recall than anticipation procedure, therefore concluding that the recall method was more sensitive to such a variable manipulation. In the present study there was no such difference between methods of learning. The two most likely explanations for this failure to replicate include the fact that Battig and Lawrence's greatest method differences were established under conditions wherein the starting point of the list was varied from trial to trial. Second, the present Method of Learning × Transfer Task interaction clearly approached significance within the contingently probed cells, suggesting that noncontingent probing was somehow interfering with acquisition. In an unpublished study by Posnansky and Battig,[3] contingent probing did not affect rate of SL at least under recall conditions, but the effects of noncontingent probing are not necessarily identical to those of the contingent probing.

---

[3] POSNANSKY, C. J., & BATTIG, W. F., Massed vs. distributed practice effects in serial learning, 1971. Copies are available from the author.

REFERENCES

BATTIG, W. F. Facilitation and interference. In E. A. Bilodeau (Ed.), *Acquisition of skill*. New York: Academic Press, 1966.

BATTIG, W. F. Advantages of recall over anticipation methods in verbal learning. *Behavior Research Methods and Instrumentation*, 1969, **1**, 217–220. (a)

BATTIG, W. F. Sequential learning and thought: An overlooked communality. In J. F. Voss (Ed.), *Approaches to thought*. Columbus, Ohio: Charles E. Merrill, 1969. (b)

BATTIG, W. F., & LAWRENCE, P. S. The greater sensitivity of the serial recall than the anticipation procedure to variations in serial order. *Journal of Experimental Psychology*, 1967, **73**, 172–178.

Battig, W. F., & Miller, S. M. Classification learning of items grouped on the basis of degree of prior learning. *Psychological Reports*, 1965, **17**, 807–814.

Battig, W. F., Brown, S. C., & Schild, M. E. Serial position and sequential associations in serial learning. *Journal of Experimental Psychology*, 1964, **67**, 449–457.

Ebenholtz, S. M. Serial learning: Position learning and sequential associations. *Journal of Experimental Psychology*, 1963, **66**, 353–362.

Keenan, V. Presentation method and learning-time constancy in serial learning. *Canadian Journal of Psychology*, 1970, **24**, 311–320.

Posnansky, C. J., Battig, W. F., & Voss, J. F. A new probe technique for identifying serial learning processes. *Behavior Research Methods and Instrumentation*, 1972, **4**, 129–132.

Postman, L. Association and performance in the analysis of verbal learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Skiba, G. S., & Battig, W. F. Effect on serial recall learning of varying levels of simultaneous presentation of parts of the list. Paper presented at the meeting of the Rocky Mountain Psychological Association, Salt Lake City, May 1970.

Voss, J. F. Serial acquisition as a function of number of successively occurring list items. *Journal of Experimental Psychology*, 1968, **78**, 456–462.

Voss, J. F. Serial acquisition as a function of stage of learning. *Journal of Experimental Psychology*, 1969, **79**, 220–225.

Woodward, A. E. Continuity between serial memory and serial learning. *Journal of Experimental Psychology*, 1970, **85**, 90–94.

Woodward, A. E., & Murdock, B. B. Positional and sequential probes in serial learning. *Canadian Journal of Psychology*, 1968, **22**, 131–138.

Young, R. K. Tests of three hypotheses about the effective stimulus in serial learning. *Journal of Experimental Psychology*, 1962, **63**, 307–313.

Young, R. K. Serial learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Young, R. K., Patterson, J., & Benson, M. M. Backward serial learning. *Journal of Verbal Learning and Verbal Behavior*, 1963, **1**, 335–338.

(Received January 29, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 194–198

DISSIPATION OF SENSORY AFTEREFFECTS AS A FUNCTION OF AXON LENGTH[1]

GEORGE SINGER[2] AND JOHN K. COLLINS[3]

Macquarie University, Eastwood, New South Wales, Australia

Sensory adaptation is explained in terms of presynaptic depletion of transmitter substance and sensory aftereffect in terms of the inefficient re-uptake mechanism. The dissipation of the aftereffect is discussed as a function of the length of axons. Using 198 *S*s in two experiments, it was shown that the visual aftereffect dissipated more rapidly than the kinesthetic aftereffect and that the recovery of kinesthetic aftereffects was a function of the length of the limb.

Sensory adaptation and aftereffect refer respectively to changes of an object property (e.g., size, orientation, color, brightness) during and after a period of constant stimulation and have been reported to occur in all sensory modalities for a wide range of object properties. The progressive change in apparent orientation of tilted edges during visual or kinesthetic stimulation is an instance of adaptation, and the apparent shift in direction of vertical or horizontal edges following stimulation is an example of an aftereffect (Gibson & Radner, 1937).

Perceptual adaptation is believed to reflect progressive diminution in the frequency of firing levels of neurons during constant stimulation, and aftereffects have been attributed to postexcitatory suppression (Barlow & Hill, 1963; Day & Singer, 1970). One difficulty with this explanation is that aftereffects outlast the supposed neural correlates.

In the experiments reported here predictions are made based on a hypothesis that the persistence of the aftereffect is understandable in terms of presynaptic depletion of transmitter substance. The conduction of the sensory impulse from the receptor to the cortex involves synaptic transmission. Continued stimulation results in a loss of transmitter since the reuptake mechanism is never 100% efficient.

Unusually prolonged and frequent stimulation of a set of neurons, selective to a specific range of stimulation (e.g., specific directions of movement, direction of tilt, color, and skin temperature), may deplete the presynaptic store of transmitters. Recently Geffen and Livett (1971) reviewed evidence for self-regulation within axon terminals of transmitter synthesis during changes in functional activity. They pointed out that increases in discharge result in an increase in the synthesis of rate-limiting enzymes necessary for the synthesis of transmitter, but that the nature and origin of the inducing signals are not known. It is probable that axoplasmic transport of either vesicles or their components is involved and the rate of vesicle travel (Geffen & Livett, 1971) is estimated to be of the order of 5–10 mm/day. Therefore, replenishing the presynaptic stores to maintain normal functioning may be a matter of hours and may depend on the length of the neurons involved.

The suggestion is that the aftereffects may be explicable by the reduction of neural activity in a period following excitation, but that the reduction in neural activity itself is dependent on a depletion of transmitter stores. Experiments within the range of short-term stimulation for 30–90 sec. have shown that stimulus duration determines the duration of the aftereffect but does not affect its magnitude (Oyama, 1953; Singer & Day, 1965).

Since axons in kinesthetic pathways are longer than those in visual pathways and since the recovery of synaptic functioning is dependent on the length of axons, the longest axon in the pathway will determine the duration of the aftereffect. Thurner (1961) reported some data which were in agreement with this hypothesis, but he used different tasks to generate the aftereffects in different modalities. In the present study a direct comparison is made of the differential recovery of visual and kinesthetic aftereffects and two kinesthetic aftereffects involving comparable tasks and equivalent amounts of stimulation.

---

[1] This work was supported by the Australian Research Grants Committee Grant A69/17176.

[2] Now at La Trobe University, Bundoora, Victoria, Australia.

[3] Requests for reprints should be sent to John K. Collins, School of Behavioral Sciences, Macquarie University, Eastwood, New South Wales, 2113, Australia.

EXPERIMENT I

The aim of the first experiment was to examine the differential recovery of the visual and kinesthetic aftereffect following equivalent amounts of stimulation. These two modalities were selected for investigation as both visual and kinesthetic aftereffects are well documented in the literature (Day & Singer, 1970), and the relatively shorter length of the optic nerve when compared with the brachial nerve makes them well suited to test the hypothesis. The tasks were comparable as *S*s were required to make judgments of the kinesthetic or visual vertical following protracted stimulation by a slanted stimulus object in one of these modalities.

Method

Subjects.—There were eight groups of 12 *S*s drawn from undergraduate students enrolled in an introductory course in psychology. They were alternately assigned to one of the four visual or four kinesthetic experimental conditions on reporting to the laboratory.

Apparatus.—The apparatus used for the visual component consisted of a light-proof box with a 90 × 2 mm. slit, the orientation of which could be controlled by both *E* and *S* by means of a 240-v. ac Philips miniature reversible motor with a long lead and manual key switch for forward and reverse movement. The motor drove a shaft on which the box was mounted housing a 6-w. fluorescent tube. On a protractor scale affixed to the mountings, the orientation of the slit could be read to the nearest .5° using a sharpened metal pointer that was aligned with the stimulus. The motor rotated in either direction at a speed of 6° per second.

The apparatus used for the kinesthetic component was similar to that used in an earlier investigation (Collins & Singer, 1968, p. 305). It consisted of a handgrip attached to a metal shaft which protruded through a wooden screen. To the rear of the shaft was attached a sharpened metal pointer which enabled readings to be taken on a 360° protractor scale to the nearest .5°. A screw clamp enabled the handgrip to be fixed at any angular position.

Procedure.—With the exception that *S* was blindfolded by use of opaque goggles on entering the laboratory for testing under kinesthetic conditions, the procedures remained the same for kinesthesis and vision. Each *S* was given a number of practice trials to familiarize him with the apparatus and to see that he understood the instructions. All *S*s were told to make their judgments as quickly as possible but not to sacrifice accuracy for speed. The *E* always set the stimulus to the objective vertical before commencing each trial. A trial consisted of a pretest, a 60-sec. stimulation period and a posttest. In the pre- and posttest, *S* adjusted the slot of light (vision) or the bar (kinesthesis) so that it seemed vertical. The four intervals between cessation of stimulation and initiation of the posttest that were used to measure the differential dissipation of the aftereffect were 0, 30, 60, and 90 sec. Following stimulation, *S* rested with occluding glasses during the delay period in vision, and with hands on knees during the delay period in kinesthesis. Each *S* was given five pretests, the mean of which was taken as his pretest score. During the stimulation period, *S*s in the visual conditions fixated the slot of light in an otherwise completely darkened laboratory and in the kinesthetic conditions gripped the slanted bar while vision was occluded. The stimulus was alternately slanted to the left or right of the vertical at an angle of 20° during this period. A single posttest was given because of the rapid initial dissipation of aftereffects (Singer & Day, 1965) and to eliminate the possible introduction of a learned effect (Collins & Singer, 1968). The difference in degrees between the pre- and posttest score served as a measure of the aftereffect. In compensating for the aftereffect, it was expected that *S*s would slant the stimulus in the same direction as the stimulating slant when making their judgments; differences in the expected direction were accordingly scored positively, while any difference in the opposite direction was scored negatively.

Results

The mean differences between pre- and posttests, together with the standard deviations, are shown in Table 1. The mean-square within-groups estimation of the population variance for vision ($MS_w = .23$) and for kinesthesis ($MS_w = 1.12$) was used to compute the confidence interval for a significant difference between the conditions ($\alpha = .05$) with 88 degrees of freedom (Rodger, 1965). Any difference between

TABLE 1

MEANS AND STANDARD DEVIATIONS OF DIFFERENCES BETWEEN PRE- AND POSTTESTS (AFTEREFFECTS) FOR THE VARIOUS CONDITIONS OF DELAY FOLLOWING STIMULATION IN VISION AND KINESTHESIS

Stimulation modality	Delay (in sec.)			
	0	30	60	90
Vision				
X̄	1.68	.74	.42	−.54
SD	.82	.29	.31	.28
Kinesthesis				
X̄	1.84	.08	1.18	.42
SD	1.13	.89	1.05	1.15

Note.—In the 0-sec. and 30-sec.-delay conditions, the difference between vision and kinesthesis was nonsignificant, and in the 60-sec. and 90-sec. conditions the difference was significant at the .05 level.

vision and kinesthesis for each of the delay periods had to exceed |.67| to be significant. The residual aftereffect in kinesthesis was significantly greater than vision for the delay conditions of 60 and 90 sec. (see Table 1).

EXPERIMENT II

A second experiment was designed to test the hypothesis intramodally. The kinesthetic modality was selected because the aftereffect in the hand (involving the brachial nerve) could be compared with the aftereffect in the foot (involving the long sciatic nerve). Many different techniques have been used to study kinesthetic aftereffects involving the upper limbs. Kinesthetic tilt aftereffects have been investigated by Collins (1967a), Day and Singer (1964), Gibson (1933), Singer and Day (1966), Zacks and Freedman (1963); kinesthetic width aftereffects by Bakan and Weiler (1963), Köhler and Dinnerstein (1947), and Moylan (1964); and kinesthetic limb displacement aftereffects by Collins (1967b), Gibson and Backlund (1963), and Nachmias (1953). Of these various techniques the one that seemed most readily adaptable to the lower limb was that of Gibson and Backlund (1963). They had *S*s place the palms of their hands at some degree of divergence or convergence during stimulation and then asked them to set palm boards parallel. Depending upon the stimulation, the later judgment was slanted in the same direction as stimulation. The same type of judgment can readily be made using the foot.

Method

Subjects.—Ninety-six male and female high school students between the ages of 14 and 15 yr. were assigned to one of eight different groups on reporting.

Apparatus.—The apparatus used to investigate the aftereffect in the hands consisted of two smooth surfaces, 20 × 12 × 1.5 cm., on which the palms of the hands could be placed. The palm boards were mounted on a wooden base and could be pivoted around their central axes. They were set 14 cm. apart at the central axes and could be adjusted to any angle of divergence or convergence by the *S*s. The angles of the planes could be read on a protractor scale to the nearest .5°.

The apparatus used to investigate the aftereffect in the foot consisted of two foot boards mounted 27 cm. apart on the top surface of a wooden stand. The foot boards were 30 cm. long and 12 cm. wide. The stand was 66 cm. wide, 38 cm. long, 27 cm. high distally from *S* and 8.5 cm. high proximally to *S*, its top surface being at an angle of 27° to the horizontal. Each foot board was connected to the stand at the middle of the heel end such that the board could be rotated around the longitudinal axis of the leg. Two straps and a wooden strip at the heel held *S*'s foot in position on the board. A metal pointer was attached to the midline of each foot board at the distal end and indicated on a protractor scale to the nearest .5° the angle of divergence or convergence of each foot.

Procedure.—Procedures during practice trials, pre- and posttests, and delay conditions were the same in all respects as for the kinesthetic component in the first experiment except that the judgment consisted of setting the hands or feet to a parallel position in the sagittal plane. During the 60-sec. stimulation period the hands or the feet were held at an angle of 20°. For half of the *S*s the limb position was one of convergence and for the other half it was one of divergence.

Results

The difference between the pre- and posttest judgments in the angles of divergence or convergence was the measure of the aftereffect. The measure was scored as positive for greater convergence after convergent stimulation, positive for greater divergence after divergent stimulation, and negative if it was in the opposite direction to stimulation.

The mean differences between pre- and posttests, together with the standard

TABLE 2

MEANS AND STANDARD DEVIATIONS OF DIFFERENCES BETWEEN PRE- AND POSTTESTS (AFTEREFFECTS) FOR THE VARIOUS CONDITIONS OF DELAY FOLLOWING STIMULATION OF THE HAND AND THE FOOT

Type of stimulation	Delay (in sec.)			
	0	30	60	90
Hand				
$\bar{X}$	3.42[a]	1.37	1.35	.14
SD	2.63	3.31	3.25	3.25
Foot				
$\bar{X}$	5.43[a]	4.18[a]	2.69	.91
SD	4.67	6.34	4.64	4.75

[a] Aftereffect significantly different from zero at .05 level.

deviations, are shown in Table 2. The mean squares within groups for the hand ($MS_w = 9.75$) and the foot (26.52) were used to test the significance of the differences of each aftereffect from zero with 44 degrees of freedom (Rodger, 1965). The aftereffect had to exceed 1.82 for the hand and 3.00 for the foot to be significant. While the aftereffect had dissipated after 30 sec. in the hand, it was still significantly different from zero in the foot (Table 2).

Discussion

The results from both experiments support the hypothesis that the duration of aftereffects is related to the length of axons involved. This not only provides a basis for the understanding of short-term effects and the paradox that the behavioral effects outlast the supposed neural correlates, but can also serve as a basis for the explanation of long-term perceptual changes such as the McCollough effect (McCollough, 1965) and the movement aftereffect reported by Masland (1969). The findings are also consistent with those reported by Thurner (1961), who observed shorter durations for visual and auditory aftereffects when compared with kinesthesis. Thurner also observed that a kinesthetic aftereffect involving the finger joints was of longer duration than a kinesthetic aftereffect involving the shoulder joint.

Although in Thurner's (1961) experiments the stimulation time was the same for all modalities, there was considerable variation in the tasks used. In the present experiments the task was uniform and, although there was a drop in the kinesthetic aftereffect in the first experiment at 30 sec., the effect was still present after 90 sec. The aftereffect at 90 sec. is in accordance with earlier work by Singer and Day (1965). The present work should be regarded as a preliminary to the study of the long-lasting perceptual effects and their duration in relation to stimulus parameters as well as axon length.

It seems possible to explain both short- and long-term aftereffects in terms of synaptic depletion, with decay in aftereffect (replenishment of presynaptic stores) depending on the period of stimulation, the length of axons, and the number of neurons involved. The explanation is confined to the underlying mechanisms of these effects which determine their duration. No attempt is made to account for the opposition between original stimulus property and apparent property. This is a different problem and explanations have been offered at the level of anatomical and functional relationships of neurons (Over, 1971).

The theory is speculative since the transmitters involved in the neural conduction of sensory impulses are largely unknown, and it needs to be shown that the type of stimulation used does in fact lead to presynaptic transmitter depletion. It may be considered that speculation is being compounded when perceptual data are accounted for by intracellular events. Nevertheless, specific long-term stimulation is consistent with transmitter depletion of nerve terminals and the long-term sensory effects are consistent with the periods of time which may be involved in repletion to restore normal synaptic functioning.

REFERENCES

BAKAN, P., & WEILER, E. Kinesthetic aftereffect and mode of exposure to the inspection stimulus. *Journal of Experimental Psychology*, 1963, **65**, 319–320.

BARLOW, H. B., & HILL, R. M. Evidence for a physiological explanation of the waterfall phenomenon and figural aftereffects. *Nature*, 1963, **200**, 1345–1347.

COLLINS, J. K. Freedom and restriction of the limb in the kinesthetic spatial aftereffect. *Australian Journal of Psychology*, 1967, **19**, 63–67. (a)

COLLINS, J. K. A kinesthetic spatial aftereffect with pronation of the forearm. *Australian Journal of Psychology*, 1967, **19**, 117–120. (b)

COLLINS, J. K., & SINGER, G. Interaction between sensory spatial aftereffects and persistence of response following behavioural compensation. *Journal of Experimental Psychology*, 1968, **77**, 301–307.

Day, R. H., & Singer, G. The relationship between the kinesthetic spatial aftereffect and variations in muscular involvement during stimulation. *Australian Journal of Psychology*, 1964, **16**, 200–208.

Day, R. H., & Singer, G. Issues in the explanation of sensory adaptation and aftereffect. Paper read at Symposium on contemporary research in psychology of perception: In Honorem Kai Von Fieandt Sexagenaru, University of Hellsenki, 1970.

Geffen, L. B., & Livett, B. G. Synaptic vesicles in sympathetic neurones. *Physiological Review*, 1971, **51**, 99–157.

Gibson, J. J. Adaptation, aftereffect, and contrast in the perception of curved lines. *Journal of Experimental Psychology*, 1933, **16**, 1–31.

Gibson J. J., & Backlund, F. A. An aftereffect in haptic space perception. *Quarterly Journal of Experimental Psychology*, 1963, **15**, 145–153.

Gibson, J. J., & Radner, M. Adaption, aftereffect, and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of Experimental Psychology*, 1937, **20**, 453–467.

Köhler, W., & Dinnerstein, D. Figural aftereffects in kinesthesis. In A. Michotte (Ed.), *Miscellanea psychologica*. Louvain, France: University of Louvain Press, 1947.

McCollough, C. Colour adaptation of edge-detectors in the human visual system. *Science*, 1965, **149**, 1115–1116.

Masland, R. H. Visual motion perception: Experimental modification. *Science*, 1969, **165**, 819–821.

Moylan, J. J. Kinesthetic figural aftereffects: Satiation or contrast. *Journal of Experimental Psychology*, 1964, **67**, 83–90.

Nachmias, J. Figural aftereffects in kinesthetic space. *American Journal of Psychology*, 1953, **66**, 609–612.

Over, R. Comparison of normalization theory and neural enhancement explanation of negative aftereffects. *Psychological Bulletin*, 1971, **75**, 225–243.

Oyama, T. Experimental studies of figural aftereffects: (1) Temporal factors. *Japanese Journal of Psychology*, 1953, **23**, 239–254.

Rodger, R. S. *Intermediate statistics*. Sydney: University Cooperative Bookshop, 1965.

Singer, G., & Day, R. H. Temporal determinants of a kinesthetic spatial aftereffect. *Journal of Experimental Psychology*, 1965, **69**, 343–348.

Singer, G., & Day, R. H. Interlimb and interjoint transfer of a kinesthetic spatial aftereffect. *Journal of Experimental Psychology*, 1966, **71**, 109–114.

Thurner, F. K. Kurzzeitige Nacheffekte unserer Wahrnehmungen. *Beiheft zur Schweiserischen Zeitschrift für Psychologie*, 1961, **41**, 1–104.

Zacks, J. L., & Freedman, S. J. Active and passive movement in the production of kinesthetic tilt aftereffects. *Perceptual and Motor Skills*, 1963, **16**, 702.

(Received June 12, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 199–200

SHORT REPORTS

EFFECT OF ACCURACY-EMPHASIZED INSTRUCTIONS ON PERFORMANCE ON AN ATTRIBUTE-IDENTIFICATION TASK

BOHDAN K. WASILEWSKI[1]

University of British Columbia, Vancouver, British Columbia, Canada

Thirty-six *S*s were randomly assigned to one of three instructional treatments (speed, accuracy emphasized, or neutral) on three problems of an attribute-identification task to test the hypotheses that the emphasis on speed: (*a*) has a detrimental effect on the performance, (*b*) is inherent in a test-like situation, and (*c*) can be reduced by emphasizing in the instructions to *S*s the detrimental effects of speed on the achievement of solution. The results supported the hypotheses.

On an attribute-identification task measured by cards to solution (CTS), Laughlin (1964) found no significant difference in performance between the speed-instructions (SI) treatment which directed *S*s to attain concepts as quickly as possible regardless of the number of card choices, and the accuracy-instructions (AI) treatment which directed *S*s to attain concepts in as few card choices as possible regardless of the speed of solution. When performance was measured by the total time to solution (TTS), the SI *S*s attained concepts significantly faster than AI *S*s. These findings are supported by Siegel's (1964) study.

The apparent noneffectiveness of AI on CTS in Laughlin's (1964) study may have been due to the failure of AI *S*s to disregard the emphasis on the speed of solution which seems to be inherent in a test-like situation. It is suggested that this inherent effect may be reduced by emphasizing in the instructions to *S*s the detrimental effects of speed on the achievement of solution.

It is hypothesized that the emphasis on speed has a detrimental effect on performance as measured by CTS and TTS. If this is true, then SI *S*s will solve an attribute-identification task with more CTS than *S*s under accuracy-emphasized instructions (AEI) and with approximately the same TTS.

It is also hypothesized that the emphasis on speed is inherent in a test-like situation. If this is true, then SI *S*s will not differ significantly in CTS and TTS from *S*s under neutral instructions (NI).

The third hypothesis is that the emphasis on speed inherent in a test-like situation can be reduced by emphasizing in the instructions to *S*s the detrimental effects of speed on the performance. If this is true, then AEI *S*s will solve the task with fewer CTS than NI *S*s and in approximately the same TTS.

Method.—The major independent variable, type of instructions (SI, AEI, or NI), was combined with three minor variables: problem (a sequence of three different problems of an attribute identification of a conjunctive two-attribute concept), set (two sets of two-binary dimensions, A and B—see the stimulus materials section), and order (3 problems presented in 3 different orders in a 3 × 3 Latin square design). Two dependent measures, CTS and TTS, were taken.

The *S*s were 36 students (18 boys and 18 girls) randomly selected from 76 volunteers out of 381 Grade 12 students at Richmond Secondary School, Richmond, British Columbia. The *S*s were randomly assigned to three treatment groups.

The stimuli were 64 white 6.35 × 8.89 cm. paperboard cards, containing all possible combinations of six binary dimensions: number of figures (one or two), size (small or large), color (red or blue), texture (solid or slashed), shape (triangular or circular), and border (solid or broken). The cards were randomly arranged in eight rows and eight columns on a 62.23 × 82.55 cm. board. The board contained 16 exemplars and 48 nonexemplars of any binary concept based on the above dimensions. A set of values of each dimension and the other set of complementary values of the six dimensions were referred to as Set A and Set B, respectively. Three conjunctive conceptual problems were constructed on the basis of randomly selected two-binary dimensions within each set. As a result, the focus card for Set A, given to 18 *S*s, had these attribute values: one, large, red, slashed, triangular figure, with a solid border; and the focus card for Set B, given to the other 18 *S*s, had their complementary values. Three conceptual problems in Set A were: one slashed figure, large triangular figure, and large figure with a solid border; and the three problems in Set B were complementary to those in Set A. Assignment to sequences was random.

Each *S* was briefed individually. The instructions consisted of two parts. In Part 1, SI *S*s were told to attain the concepts as quickly as possible, regardless of the number of card choices; AEI *S*s were told to attain the concepts with as few card choices as possible and to disregard the time, since the speed of performance would prevent them from achieving solution with the minimum of card choices; and NI *S*s were told only to attain the concepts. Part 2 contained general learning instructions including a practice identification of all exemplars of the concept "large red figure," the information that all concepts

---

[1] Requests for reprints should be sent to Bohdan K. Wasilewski, Faculty of Education, University of Victoria, P. O. Box 1700, Victoria, British Columbia, Canada.

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR CARDS TO SOLUTION (CTS) AND TOTAL TIME TO SOLUTION (TTS)

Performance and treatment	$\bar{X}$	SD
CTS		
SI	16.53	8.29
NI	14.92	7.51
AEI	7.22	1.89
TTS		
SI	277.3	166.2
NI	245.5	120.6
AEI	345.7	187.9

Note.—n = 12 in each treatment condition. All means and standard deviations are given in seconds.

to be identified were composed of two out of six dimensions represented on each stimulus card, and the instructions that after each card selection *S* could offer one hypothesis, which then would be identified by *E* as "correct" or "incorrect." After Parts 1 and 2 had been read to *S*, *E* answered questions by rereading the appropriate part of the instructions; and *S* was given the first problem and the focus card (located in the top left corner of the board for Set A and in the top right corner for Set B). Choice of each stimulus card was made by *S* at his own rate. Upon selection of each card, *E* designated whether it was an exemplar or nonexemplar. When *S* offered the correct hypothesis, he proceeded without a pause to the next problem, until he solved all three problems.

Results and discussion.—The effects due to the three minor variables (problem, set, and order) were nonsignificant, and consequently, the results were collapsed over these variables for the analysis of the major variable. The means and *SD*s for the observed responses in terms of CTS and TTS are shown in Table 1.

The main effect due to the instructional variable in terms of CTS was significant, $F\ (2, 33) = 6.93$, $p < .05\ (p = .003)$. The Newman-Keuls comparison between ordered means was nonsignificant for SI–NI ($p = .05$), and significant for SI–AEI and NI–AEI ($p = .01$). The main effect due to the instructional variable in terms of TTS was nonsignificant, $F\ (2, 33) = 1.42$, $p > .05\ (p = .255)$.

The results were as hypothesized. The fact that the performance of SI and NI *Ss* seemed to be similar indicates that the emphasis on speed is inherent in a test-like situation. The fact that AEI *Ss* solved the task with fewer CTS, and within TTS not significantly different from TTS of SI *Ss*, indicates that speed has a detrimental effect on performance.

The results from the present study are the reverse of the findings of Laughlin (1964). Whereas in the Laughlin study AI *Ss* solved the task with no significant difference in CTS and with larger TTS than SI *Ss*, in the present study AEI *Ss* solved the task with fewer CTS and with no significant difference in TTS than SI *Ss*. Thus it appears that AEI *Ss* do perform more efficiently than AI *Ss* in terms of CTS and TTS. This may be due to the fact that AI *Ss* were simply told to disregard time, while AEI *Ss* were given the reason for disregarding time. Telling AEI *Ss* that speed would have a detrimental effect on achievement of the solution apparently reduced the emphasis on speed which seems to be inherent in the test-like situation.

REFERENCES

LAUGHLIN, P. R. Speed versus minimum-choice instructions in concept attainment. *Journal of Experimental Psychology*, 1964, **67**, 596.

SIEGEL, P. S. Concept formation's response-time considerations. *Psychological Reports*, 1964, **14**, 435-442.

(Received November 10, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 201–203

RE-PAIRING, NUMBER OF ALTERNATIVES, AND MEANINGFULNESS IN VERBAL DISCRIMINATION LEARNING[1]

JAMES R. ULLRICH[2]

University of Montana

The effect of the meaningfulness of the alternatives (words vs. trigrams), the number of items (2 vs. 4), and the presence or absence of a re-pairing procedure (constant vs. variable procedure) on verbal discrimination (VD) learning were investigated. Analyses of the total number of errors indicated that the re-pairing procedure had no effect and that there existed an interaction between the number of alternatives and meaningfulness. The precise direction of this interaction was a function of a correction for differential guessing rates. The results were interpreted as supporting a frequency theory of VD learning.

Ekstrand, Wallace, and Underwood (1966) have proposed a frequency theory of verbal discrimination (VD) learning in which performance is assumed to be a function of the differential accrual of frequency units to the right (R) and the wrong (W) items in a VD task. Kanak and Dean (1969) found a decrement in List 2 performance of a VD transfer task if List 2 is constructed from List 1 by re-pairing the R and W items; this decrement was not predicted by frequency theory since the re-pairing operation does not alter the frequency mechanisms which are assumed to be the basis for discrimination. In order to account for the decrement observed in the transfer task, Wallace and Nappe (1971) liberalized the counting postulate of frequency theory by appealing to notions of variability and yoking. The present investigation was designed to determine if it is necessary to modify the counting postulate for single-list VD studies and to determine the role of incidentally learned intrapair associations by comparing the performance of a normal VD group with the performance of a variable group in which the R and W items were randomly re-paired on each trial. In an unpublished paper, Lovelace (1968) has suggested that this procedure will have a negligible effect upon performance.

In addition, frequency theory may be interpreted as predicting an interaction between the number of alternatives in the VD list and the ease with which frequency units are attached to the W items. To explain this interpretation, first consider the case where the items in the VD task are common words, that is, where representational responses are easily made by *S*. Four-alternative (4A) lists should be easier to learn than two-alternative (2A) lists because performance is posited to be a function of the difference or the ratio between the frequency units attached to the R and the W items and because the same number of frequency units are assumed to be spread over three W words in the 4A lists as opposed to one W word in the 2A lists. This dispersal of frequency units results in fewer frequency units attached to each W word; hence, there is a greater difference or ratio and faster acquisition for the 4A lists. Supporting this prediction, Radtke and Foxman (1969) have concluded that 4A lists of words were easier to learn than 2A lists.

Next, consider the case where the stimulus materials are composed of trigrams. Ekstrand et al. (1966, p. 576) have noted that the representational responses might be formed to the nonintegrated individual letters as opposed to the trigrams as a unit; if this is so, it follows that the representational responses to the trigrams would be less reliably formed because of confusion between letters. Until the verbal units are sufficiently integrated, frequency could not be expected to be utilized as a cue. Since integration would be slower with a greater number of alternatives, it would be predicted that 2A lists of trigrams would be easier to learn than 4A lists. This reversal of predictions for the 2A and the 4A lists between words and trigrams forms the basis for predicting the interaction between meaningfulness and number of alternatives. It should be noted that it is extremely difficult to specify the precise directionality of these predictions because of the different initial guessing rates of the 2A and of the 4A lists. Any attempt to correct for guessing requires one to make rather strong assumptions about the underlying mathematical structure of the theory. The inclusion of a correction factor would not affect the prediction of an interaction but would easily affect the direction of predictions for the 2A and the 4A lists. Ekstrand et al. have also noted that frequency theory makes no specific predictions about the relative difficulty of VD lists that vary in meaningfulness. Frequency theory with the original counting postulate predicts that there would be no interactions between the presence or absence of re-pairing and other variables since the frequency mechanisms would not be altered by the re-pairing procedure.

Method.—One hundred and thirty-six psychology students served as *Ss* as part of course requirements. None of the *Ss* had any formal classroom instruction in verbal learning or had participated in verbal learning experiments.

This study was a 2 × 2 × 2 completely randomized factorial design in which the meaningfulness of the items (words vs. trigrams), the number of alter-

---

[1] This research was supported by Grant MH18695-01 from the National Institute of Mental Health, United States Public Health Service.

[2] Requests for reprints should be sent to James R. Ullrich, Department of Psychology, University of Montana, Missoula, Montana 59801.

natives (two vs. four), and the pairing procedure (constant vs. variable) were permuted. The eight groups were denoted W2C, W2V, W4C, W4V, T2C, T2V, T4C, and T4V. For the 2A conditions, constant pairing referred to the general VD procedure in which each R item was paired with one W item, and this pairing was maintained throughout the experiment. Under the variable-pairing condition, each R item was re-paired with a different W item on each successive trial. Under the 4A conditions, quadruples were formed by randomly selecting three W items to be presented with each R item. For the constant-pairing condition, the same quadruples were presented on every trial; under the variable-pairing conditions, a different set of quadruples was formed and presented on each successive trial.

Each list was constructed from one of two 64-item pools. The pool of words was formed by random selection from those five-letter words which had a Thorndike-Lorge frequency rating of AA; the pool of trigrams was formed by a random selection of those trigrams from Archer's (1960) list, which had meaningfulness ratings between 48 and 55.

The R items for the word and for the trigram conditions were chosen by randomly selecting 16 items from each pool. An additional sample of 16 formed the W items for the two-alternative conditions; the remaining 32 items in each pool served as additional W items for the 4A lists.

The W4C list was formed by randomly creating 16 quadruples consisting of 1 R word, 1 W word from the initial 16, and 2 W words from the remaining 32 items of the pool of words. The order of presentation between and within quadruples was randomly determined for each of 30 trials. No attempt was made to counterbalance order within or between quadruples. The W4V list was formed from the W4C list by randomly interchanging the W words on each trial. The R items occupied the same position within and between quadruples for each trial as in the W4C list. The W2C list was constructed from the W4C list by deleting the 32 W items mentioned in the previous paragraph. Within each pair of the W2C list, the order of the R and W items was identical to the order of the same imbedded pair in the W4C list. The W2V list was formed from the W2C list by randomly interchanging the W items on each trial. The trigram lists were constructed in the same fashion using the same randomizations as the word lists.

An IBM 029 key punch was used to present automatically the stimuli, record the responses, and provide feedback. The pairs or quadruples were punched, printed, and left adjusted on a Hollerith data card beginning in column 3. One blank column separated the individual items. The R item of each set was punched, printed, and left adjusted beginning in column 75 of the same card. Sixteen cards formed one trial, and a blank card separated the trials; the entire set of stimulus materials consisted of a deck of $(16 + 1) \times 30 = 510$ cards for each of the eight experimental conditions. The *S*s made their responses by punching, without printing, a number indicating the position of the item in the set they chose as correct. Each *S*'s responses were recorded in a column between 41 and 65; the particular column used for a given *S* was dependent upon the order of arrival for the experiment and was constant across groups. The card feed bed and the card read bed of the key punch were partially covered to prevent *S* from seeing other *S*'s responses or from seeing the feedback before a choice was made. At any moment each *S* could see only the particular set of items from which a choice was to be made and the feedback from the previous set of items. The keyboard was partially covered to allow only the responses 1 and 2 for the two-alternative lists and 1, 2, 3, and 4 for the four-alternative lists.

Each *S* was assigned to a group based on the order of appearance via a randomized block procedure and was given instructions appropriate to his group. The instructions differed only with respect to the examples used to illustrate the construction of the list for his group. The *S* was then given 6 trials on a practice list consisting of four sets of items. If he did not obtain a criterion of 2 successive trials without errors, *S* was given an additional 6 trials on the practice list. The construction of the practice lists differed from the construction of the experimental lists only with respect to the number of sets of items within the list. When *S* finished the practice list, any questions were answered and the main experiment began. Each *S* proceeded through all 30 trials at his own pace.

Results and discussion.—The mean number of errors made on Trials 2–30 for each treatment condition are presented in Table 1. A $2 \times 2 \times 2$ analysis of variance indicated that meaningfulness and the Meaningfulness × Number of Alternatives interaction were significant, $F\ (1, 128) = 8.77$ and 7.34, respectively, $p < .01$. Pairing procedure, $F\ (1, 128) = 1.67$, number of alternatives, $F\ (1, 128) = 2.47$, the interaction of the number of alternatives and pairing procedure, $F\ (1, 128) = 2.00$, the interaction of meaningfulness and pairing procedure, $F\ (1, 128) = 1.65$, and the triple interaction, $F\ (1, 128) = .010$, were not significant, thus indicating that the re-pairing procedure had a minimal effect upon performance. The mean number of errors for W2, W4, T2, and T4 were found to be 39.3, 34.3, 40.4, and 58.8, respectively, by collapsing over the pairing procedures. Individual comparisons among these means revealed that all significant effects were due to the greater number of errors under the T4 condi-

TABLE 1

MEAN NUMBER OF ERRORS FOR EACH TREATMENT CONDITION

Pairing procedure	No. of alternatives and meaningfulness			
	Two		Four	
	Words	Trigrams	Words	Trigrams
Constant	36.4	31.5	37.2	56.5
Variable	42.1	49.2	31.5	61.1

tion. The lack of significant differences between W2 and W4 is consistent with results obtained by Radtke and Foxman (1969), who found that a four-alternative word list was easier only when the number of errors was corrected for differences in guessing probability. The present results indicated equality of W2 and W4 on uncorrected total errors; correction for differences in guessing probability would show that W4 was easier than W2. Because of the strong assumptions involved in determining the appropriate correction factor, no attempt was made to show differences between W2 and W4. It should be noted, however, that the significance of the interaction between number of alternatives and meaningfulness would not be affected by a correction factor.

The lack of significant differences due to pairing procedures and the significance of the interaction between meaningfulness and number of alternatives is in accord with the original statement of frequency theory. This is somewhat surprising in the sense that *S*s in the constant groups undoubtedly possessed information in the form of incidentally learned intrapair associations which did not improve performance. The fact that the pairing procedure was not a potent variable is further supported by the unpublished work of Lovelace (1968) and by an investigation of the statistical power of the analysis of variance. The power of this test was sufficiently great to detect a difference of one error per item per *S* with a probability of approximately .95.

REFERENCES

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74**(10, Whole No. 497).

EKSTRAND, B. R., WALLACE, W. P., & UNDERWOOD, B. J. A frequency theory of verbal discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

KANAK, N. J., & DEAN, M. F. Transfer mechanisms in verbal discrimination. *Journal of Experimental Psychology*, 1969, **79**, 300–307.

LOVELACE, E. A. Verbal discrimination: Meaningfulness and re-pairing. Paper presented at the meeting of the Eastern Psychological Association, Washington, D. C., April 1968.

RADTKE, R. C., & FOXMAN, J. Number of alternatives and similarity in verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 537–544.

WALLACE, W. P., & NAPPE, G. W. Re-pairing "rights" and "wrongs" in verbal discrimination learning. *Journal of Experimental Psychology*, 1971, **87**, 355–360.

(Received November 22, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 203–205

EFFECT OF FORWARD HEAD INCLINATION ON VISUAL ORIENTATION DURING LATERAL BODY TILT

N. J. WADE[1]

Max-Planck-Institut für Verhaltensphysiologie, Seewiesen, West Germany

Judgments of the visual vertical were made during lateral body tilts up to 180°, for different degrees of head inclination in the sagittal plane. The deviations of the visual vertical from the body median plane were not influenced by the forward head inclination at any body tilt. It was concluded that the lateral shear force on the utricular statoliths is the adequate stimulus for visual orientation in the frontal plane.

The effect of body tilt on visual orientation is now well documented. In the absence of any visual framework a line of light can be adjusted relatively closely to the gravitational vertical during tilt. There is, however, a general function relating visual orientation to the degree of body tilt. For tilts of less than about 60° the visual vertical tends to be on the opposite side of the gravitational vertical to tilt, whereas the reverse occurs for greater tilts up to 180°; in this last position, the judgments are similar to those with the body upright (see Day & Wade, 1969).

The relative proximity of the visual and gravitational verticals during body tilt may be interpreted in terms of visual orientation constancy (Day & Wade, 1969). The mechanism most commonly implicated in the maintenance of constancy is that of the statolith system and, more particularly, the utricular statolith system. The effective stimulus for the utricular system is the force acting tangentially to the macular surface (Trincker, 1962). This force can be modified by tilt or centrifugation, or a combination of the two, and it is from these operations that the evidence for utricular involvement derives (see Day & Wade, 1969).

Little attention has been directed to factors within the utricular system which may influence visual orientation. For example, the position of the head in the sagittal plane determines the direction of action of the force along the macular surface. Consideration has been given only to the lateral component of this force which, because of the orientation of the maculae, is essentially equivalent to the force acting at right angles to the head median plane. The present experiment was designed to examine whether the lateral component alone is of importance for visual orientation in the

---

[1] Requests for reprints should be sent to N. J. Wade, who is now at the Department of Psychology, University of Dundee, Dundee DD1 4HN, Scotland.

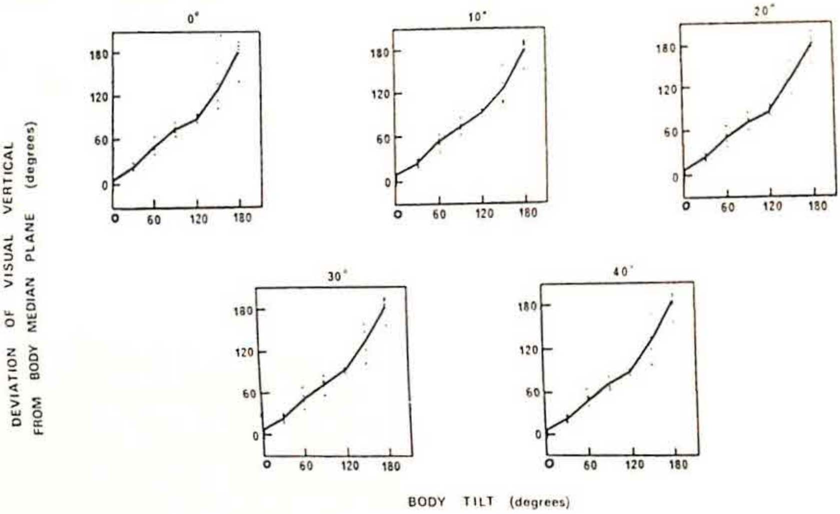


FIG. 1. Deviation of the visual vertical from the body median plane as a function of lateral body tilt. (A diagonal line would indicate alignment of the visual vertical with the direction of gravity. The separate graphs represent the data for each head inclination from upright (0°) to 40° forward in the sagittal plane. The points signify the averaged data for each *S*, and the lines join the overall means.)

frontal plane. To this end, visual orientation was measured during lateral body tilts involving different degrees of head inclination in the sagittal plane. When the head is upright (here defined as that position in which the teeth are impressed in a horizontal biteboard), the line joining the base of the orbit to the external auditory meatus is tilted backward by about 12°, such that the utricular maculae would be inclined backward by approximately 40° from the horizontal (see Wade & Schöne, 1971). Conversely, with the head inclined forward by 40° the plane of the utricular maculae would be in the region of the horizontal. The consequences of lateral body tilt in these two head positions are different magnitudes and directions of the force operating tangentially to the macular surfaces; these forces can be resolved into the components acting laterally and backward or forward. In the former case, with the head upright there is a backward-acting force; lateral tilt reduces the magnitude of this backward vector until at 90° it is 0, and the lateral vector alone is operating with a maximum value of 1 g. Tilts beyond 90° decrease the lateral component, but the additional vector is directed forward, reaching its maximum at 180° tilt. In the latter instance, with the head inclined 40° forward so that the utricular maculae are about horizontal, the only changing component is the lateral one, which follows the same function as in the first case, i.e., increasing from 0 at 0° (upright) to 1 g at 90° and subsequently decreasing to 0 again at 180° tilt. In addition to the two head positions mentioned above three intermediate inclinations were also examined. For all head inclinations the value of the lateral vector would have been the same at any given tilt, but the backward–forward vector would have differed for all tilts other than 90°, where the lateral vector alone is operating. If the backward–forward component of the macular force does affect visual orientation in the frontal plane, this would be detected by the head inclination by body tilt interaction, due to the reversal in the direction of this component for tilts beyond 90°.

Method.—The apparatus has been described in detail elsewhere (Schöne, 1964; Udo de Haes, 1970). It consisted essentially of a framework upon which *S* lay prone. Adjustable side and top supports minimized any trunk movements during tilt. The head position was controlled by means of an individual dental-composition biteboard, side head clamps, and a head strap. The orientation of the head in the sagittal plane was varied by means of the adjustable biteboard, enabling different degrees of forward head inclination to be studied while maintaining the same trunk position. The visual stimulus object was a line 10 mm. long and 1 mm. wide, located 25 cm. from *S*'s eyes and with its center at eye level. The orientation of the line could be controlled by *S* using a knob by his right hand. The plane of rotation of the line was constant throughout; it was vertical and in the frontoparallel plane when the head was upright. The apparatus was situated in a dark room.

Four *S*s, three female and one male, participated in the experiment. All were well practiced in making visual orientation judgments under conditions similar to those involved in this experiment.

The five head inclinations investigated were 0° (biteboard horizontal) and 10°, 20°, 30°, and 40° forward. Lateral body tilts of 0° (upright) and 30°, 60°, 90°, 120°, 150°, and 180° right were studied for each head position. Within a session, *S* made four adjustments to the visual vertical at each lateral tilt in ascending order. The body was rotated at 2°/sec between the tilt positions. Each *S* underwent 20 sessions. The order of the forward head inclinations was randomized for the first 5 sessions, one for each inclination, and this order was reversed successively for the remaining blocks of 5 sessions.

Results and discussion.—The data were averaged over sessions and *S*s for each body tilt and head inclination. The deviations of the visual vertical from the body median plane are shown in Fig. 1; the curves for each head inclination are given separately. A diagonal line would indicate perfect constancy; i.e., the visual vertical would be displaced from the body median plane by the same degree as body tilt, with consequent correspondence between the visual and gravitational verticals. Under all tilt conditions the visual vertical deviated from the gravitational vertical in the direction of body tilt (termed the A effect). Usually for tilts of 30° the opposite (called the E effect) occurs. However, the absence of the E effect in the present experiment may have been due in part to the lying posture of *S*, and also to making a series of judgments in the tilted position. The A effect increased with increasing lateral tilts up to 120°, whereupon it decreased until the vertical was located relatively accurately (on average) at 180°. There were, however, large individual differences at the extreme tilts, as can be seen from the graphs in Fig. 1. These results are in general agreement with those reported by Udo de Haes (1970), although the individual variations were still large at 180°.

The degree of forward head inclination did not significantly influence visual orientation, $F\ (4,\ 12) = .70,\ p > .05$, nor was there any significant interaction between head inclination and lateral body tilt, $F\ (24,\ 72) = 1.32,\ p > .05$.

The results, therefore, demonstrate quite clearly that it is the lateral utricular shear component alone that is relevant for visual orientation in the frontal plane. The variation in the magnitude of the other component, whether backward acting for tilts less than 90° or forward acting for greater tilts, has no systematic effect on the visual vertical. The backward-acting component, however, does affect visual orientation judgments in the sagittal plane (Ebenholtz, 1970). That is, with variation in forward head inclination the line appears to tilt in the sagittal plane. This effect was not investigated in the present experiment as the light line remained in a fixed relation to *S*'s trunk throughout. Furthermore, the different head inclinations were associated with different gaze directions, varying from straight ahead to 40° elevation. Such elevation appears to be accompanied by slight ocular torsion, but this is likely to have been minimal under the binocular viewing conditions employed, due to the cyclofusional eye movements generated by the stereoscopic cues from the line. No systematic shifts in the visual vertical were evident for the different head inclinations with the body upright.

It would have been instructive to examine head inclinations backward of the upright, but constraints of the apparatus did not permit this: With the trunk prone the most extreme head position possible corresponded to that with the biteboard horizontal. Nonetheless, it is reasonable to assume that the results obtained for forward head inclinations would apply in like manner to backward inclinations.

The foregoing analysis has considered only the influence of the utricular statoliths on judgments of the visual vertical. While it is known that mechanoreceptors in the trunk do influence visual orientation (Day & Wade, 1969), these would have been equivalently stimulated for all the head positions studied. The changes in head inclination would have resulted in a difference in functioning of the saccular statoliths. However, there is a paucity of evidence concerning the function of the saccular system in general and little, if any, implicating it directly in visual orientation. Therefore, at this initial level of analysis it appears valid to consider orientation constancy primarily in terms of the utricular statolith system.

In conclusion, the results of the above experiment indicate that visual orientation in the frontal plane is a function of the lateral shear component operating on the utricular statolith system. The question now arises as to the precise relationship between lateral utricular shear and visual orientation (cf. Schöne, 1964).

REFERENCES

Day, R. H., & Wade, N. J. Mechanisms involved in visual orientation constancy. *Psychological Bulletin*, 1969, **71**, 33–42.

Ebenholtz, S. M. Perception of the vertical with body tilt in the median plane. *Journal of Experimental Psychology*, 1970, **83**, 1–6.

Schöne, H. On the role of gravity in human spatial orientation. *Aerospace Medicine*, 1964, **35**, 764–772.

Trincker, D. The transformation of mechanical stimulus into nervous excitation by the labyrinthine receptors. *Symposium of the Society for Experimental Biology*, 1962, **16**, 289–314.

Udo de Haes, H. Stability of apparent vertical and ocular countertorsion as a function of lateral tilt. *Perception & Psychophysics*, 1970, **8**, 137–142.

Wade, N. J., & Schöne, H. The influence of force magnitude on the perception of body position. I. Effect of head posture. *British Journal of Psychology*, 1971, **62**, 157–163.

(Received December 10, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 206–208

NATURAL WORDS AS PHYSIOLOGICAL CONDITIONED STIMULI: FOOD-WORD-ELICITED SALIVATION AND DEPRIVATION EFFECTS[1]

ARTHUR W. STAATS[2] AND ORMOND W. HAMMOND

University of Hawaii

The *S*s salivated more to auditorily presented food words than to nonfood words. Food-deprived *S*s salivated more to food words than did nondeprived *S*s. The results lend support to a classical conditioning theory of the emotion- (attitude) eliciting function of words.

A classical conditioning theory of the emotional (attitudinal) function of language (as well as sensory or image word meaning) suggests that certain words are systematically paired with environmental events that elicit physiological responses. The words come to elicit part of the response elicited by the environmental event with which the word is paired, and the word can transfer this conditioned response to other stimuli with which it is later paired (see Staats, 1968a, 1971).

The above publications describe experiments supporting the first-order conditioning of emotional responses to words, as well as experiments demonstrating the first author's higher order conditioning procedure, using words to elicit the emotional response. The latter experiments have not been unanimously accepted as demonstrations of classical conditioning, with some experiments supporting the conditioning approach and others contesting it. There is need for further study. Moreover, the internal demands of the theory construction task require further empirical specification. One essential requirement is direct empirical demonstration of the relationship of the conditioning principles to the physiological level of observation assumed in the theory.

The theory specifies physiological emotional response elicitation by appropriate words in one's language. One example, a basic hypothesis of the present study, is that food words are systematically (if not always) paired with food, and should come to be conditioned stimuli for the salivary response.[3] In the study to be described, the expected salivary-response-eliciting function of a sample of food words was tested, plus another principle from conditioning theory. Finch (1938) first showed that extent of the conditioned (as well as unconditioned) salivary response increases with food deprivation up to 72 hr. The present study tested the deprivation principle with humans, on the extent of salivation to food words.

Method.—Twenty-two male and 21 female undergraduates volunteered for the study; 3 were excluded for deviating conditions. The overall data matrix was a $2 \times 2 \times 2 \times 2 \times 5$, or $S_{40}(A_2B_2C_2)D_2E_5$ design, where the last two factors were repeated measures. The *S*s were nested within A (sex: M or F), B (deprivation: deprived or nondeprived), and C (word list: 1 or 2). The D variable was the 2 types of words: food words and nonfood words. E was the number of trials in each type of word.

Each word list had five nonfood words and five food words, randomly mixed. The two lists were as follows: Word List 1, BLACK, PANCAKES, ROAST BEEF, SECOND, BACON, SQUARE, ANVIL, FRENCH FRIES, ICE CREAM, IRONY; Word List 2, DARK, EGGS, ADVICE, HAMBURGER, LOBSTER, FARCE, EDGE, RICE, GRAVY, HEAT.

The *S*s to be deprived were asked to not eat breakfast and lunch (solid food) on the day of the experiment, supposedly to assure uniformity among *S*s in an experiment involving digestive processes. Nondeprived *S*s were asked to eat lunch before reporting for the experiment (which was conducted between 12 noon and 2:30 p.m. for all *S*s). A drink of water was available prior to the experiment. The *S*s were told the study involved collection of saliva for later analysis since different words could influence the digestive processes. Each *S* followed instructions to swallow all saliva prior to each trial, a cotton swab was inserted under *S*'s tongue, a word was presented auditorily, and 20 sec. later the cotton was removed. The *S*s were told to repeat the word to themselves once every several seconds. Following Razran's (1936) technique, the cotton swabs were weighed before and after absorbing *S*'s saliva during each trial, the weight difference being the measure of amount of salivation. Five adaptation trials were run prior to the first word presentation. At the end of the data collection *S*s answered certain questions, for example, time since last eating.

A Mettler, Type P162, electronic chemical scale, accurate to .001 gm., was employed for rapid weighing of the saliva-laden cotton swabs. The swabs were Johnson and Johnson, No. 3., 1.5 in. long. A partition separated *S*s from view of *E*'s materials.

Results.—The results of the ANOVA of the salivation scores revealed no differences over all the specific words due to sex (A), deprivation (B), or trials (E). A central finding was that all *S*s across all conditions did salivate more to food words than to nonfood words ($F = 20.57$, $p < .001$). Although the word lists were composed by random selection of words, the word-list (C) variable produced significantly different amounts of salivation ($F = 5.63$, $p < .025$). Salivation was greater to List 1 than to 2. Lack of significance in the relevant higher order interaction terms involving the major variables of deprivation and word condition (BC, CD, BCD) indicated that the word-list difference did not interfere with the other two effects.

---

[1] The authors are grateful to Karl A. Minke, who provided advice on the experimental design and statistical analysis.

[2] Requests for reprints should be sent to Arthur W. Staats, Department of Psychology, University of Hawaii, 2430 Campus Road, Honolulu, Hawaii 96822.

[3] Razran (1936) measured the amount of his own salivation when he said one word, "saliva," to himself in seven languages with which he was differently familiar—with salivation occurring to the extent of that familiarity. The use of only one *S* limits generalization, however, and the use of oneself introduces other limitations in interpretation. The present theoretical analysis, in specifying the manner in which a verbal stimulus comes to elicit an emotional response, also hypothecates that a *class* of food words will elicit salivation.

The *Ss'* answers concerning their food consumption prior to the experiment revealed that the deprived group indicated a mean of 15.8 hr. of food deprivation ($SD = 1.91$) and the nondeprived group had a mean of 2.5 hr. of deprivation ($SD = 1.77$). Figure 1 depicts the effects of the food deprivation on the extent to which *Ss* salivated to the food words as compared to their salivation to the nonfood words. Both the deprived and the nondeprived *Ss* salivated similarly to the nonfood words. The nondeprived mean salivation was .0661 gm., the deprived *Ss'* mean was .0657 gm. Both groups salivated more to the food words. The nondeprived *Ss* have a mean of .0756 gm. The effects of food deprivation, however, are evident in the greater salivation of the deprived group, whose mean salivation is .1011.

The deprivation effects, as revealed in the BD (Deprivation × Word Condition) interaction, is significant ($F = 6.80$, $p < .025$). No significant difference occurred in amount of salivation between the deprived and the nondeprived *Ss* in response to the nonfood words. While both groups salivated more to the food words, deprived *Ss* salivated significantly more to these words than did the nondeprived *Ss* ($t = 2.16$, $p < .025$). The only other significant term in the overall analysis of variance involved the four-way Sex × Deprivation × Word Condition × Trials interaction, $F = 2.90$, $p < .025$, again suggesting a differential word effect. The data did not meet the assumption of homogeneity of variance. However, the other assumptions are met by the data, and the present levels of significance plus the robust nature of the F test in the face of homogeneity violations make any qualification of the results unnecessary.

The results of *Ss'* hunger rating suggested that the deprivation manipulations produced different experienced hunger. That is, the deprived *Ss* rated themselves at a mean of 4.70 on the 7-point Full–Hungry scale. The nondeprived had a mean of 2.55.

Discussion.—The results showed that food words elicited a salivary response, and that the effect was affected by the food-deprivation operation. The latter is a direct implication of basic classical conditioning principles. The study establishes the necessary link to the physiological level of responding for this class of affective words. As such it helps support the general classical conditioning theory of the emotional characteristics of language that has been elaborated.

It should be noted that this is an overlapping theoretical structure in which various principles and experimental results are included (Staats, 1968a, 1971). (*a*) For example, there is evidence that physiological responses can be conditioned to language stimuli and that semantic and attitudinal (evaluative) ratings of these stimuli are changed by the conditioning. (*b*) It has been demonstrated that natural words that elicit emotional or attitudinal responses, as shown by ratings, also have reinforcement value, and can strengthen instrumental responses on which they are contingent. (*c*) Higher order classical conditioning appears to occur when emotion-eliciting words are paired with new conditioned stimuli, as measured by rating scales. (*d*) Moreover, this higher order conditioning of emotional responses—when food words are used as the unconditioned stimulus (UCS) words—shows an enhanced effect when *Ss* have been deprived of food (Staats, Minke, Martin, & Higa, 1972). (*e*) The present study shows the same effect of deprivation, directly, with the physiological response of salivation. When these and other analyses and experimental results are considered together, there appears to be a theoretical and empirical structure with a solid and growing foundation.

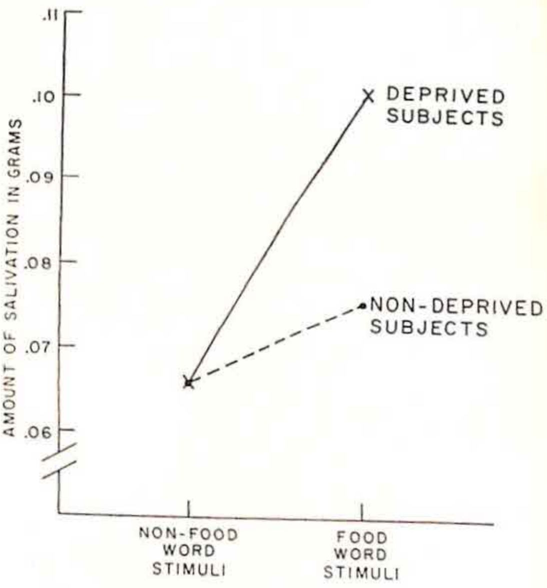


FIG. 1. Mean salivation to nonfood words and food words by nondeprived and deprived *Ss*.

Moreover, the classical conditioning theory of the emotional aspects of language has implications that extend to various areas of study (see Berkowitz, in press; Byrne & Clore, 1970; Staats, 1968b). For example, Hekmat and Vanian (1971), using the present author's language conditioning methods, paired the word SNAKE with positive emotional words. It was found that the phobic *Ss* so treated changed their attitudinal ratings of the word SNAKE. Moreover, direct measurement of behavior indicated that these *Ss* would approach and touch an actual snake more than would control *Ss*. Present applications of the conditioning principles and procedures should be bolstered by additional basic studies.

REFERENCES

BERKOWITZ, L. Words and symbols as stimuli to aggressive responses. In J. F. Knutson (Ed.), *Control of aggression.* Chicago: Aldine, in press.

BYRNE, D., & CLORE, G. L. A reinforcement model of evaluative responses. *Personality: An International Journal*, 1970, **1**, 103–128.

FINCH, G. Hunger as a determinant of conditional and unconditional salivary response magnitude. *American Journal of Physiology*, 1938, **123**, 379–382.

HEKMAT, H., & VANIAN, D. Behavior modification through covert semantic desensitization. *Journal of Consulting and Clinical Psychology*, 1971, **36**, 248–251.

RAZRAN, G. H. S. Salivating and thinking in different languages. *Journal of Psychology*, 1936, **1**, 145–151.

STAATS, A. W. *Learning, language, and cognition.* New York: Holt, Rinehart & Winston, 1968. (a)

STAATS, A. W. Social behaviorism and human motivation: Principles of the attitude-reinforcer-discriminative system. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes.* New York: Academic Press, 1968. (b)

STAATS, A. W. Linguistic-mentalistic theory versus an explanatory S-R learning theory of language development. In D. I. Slobin (Ed.), *The ontogenesis of grammar: A theoretical symposium.* New York: Academic Press, 1971.

STAATS, A. W., MINKE, K. A., MARTIN, C. H., & HIGA, W. R. Deprivation–satiation and strength of attitude conditioning: A test of altitude-reinforcer-discrimination theory. *Journal of Personality and Social Psychology*, 1972, **24**, 178–185.

(Received December 20, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 208–210

PICTURE VERSUS WORD AND RELEVANT VALUE "RELATEDNESS" IN RULE-LEARNING PROBLEMS[1]

A. KEITH BARTON[2]

University of Texas at Austin

The present report involves an application of two verbal learning variables in the area of *rule-learning* problems. More specifically, the role of prior association of relevant values and imagery is investigated. In addition to rule complexity differences, picture presentation was found superior to word presentation. Contrary to previous expectations, manipulation of relevant value relatedness produced no discernible saliency advantages, and conditional rules presented the most difficult learning task. Results were discussed in terms of retrieval advantages for pictures, possible unit processing of the two relevant values, and stimulus sequencing effects on rule acquisition. A major intent of the present report was to focus attention on other variables in addition to those currently being studied in rule learning.

A reliable finding within the rule-learning (RL) literature involves an ordered difficulty in rule application to concept problems (Bourne, 1970; Haygood & Bourne, 1965). Conjunctive (CJ) and disjunctive (DJ) rules are easier to acquire than conditional (CD) and biconditional (BCD) rules. Several explanations have already been reviewed for expected and actual rule-learning differences (Bourne, 1970) and will not be discussed in the present report.

Using geometric stimuli, Bourne (1970) has shown a CJ $<$ DJ $<$ CD $<$ BCD ordering in rule difficulty (read "$<$" as easier). Four dimensions with three values per dimension were employed with color and form as the relevant dimensions (RDs). However, the RDs can be described as neutral in that "red" and "square" have little a priori association value (AV).

The present article involves the application of two verbal learning variables—namely, associative strength and picture versus word manipulations, to *rule-learning* problems in concept learning.[3] Consequently, RDs which were thought to have a high a priori color–object AV were employed—e.g., "red–firetruck" as the relevant values (RVs). The AV manipulation should increase the saliency of the RVs, thereby producing a greater tendency to ignore the irrelevant attribute values. Such a manipulation should lead to improved performance in RL problems.

Word versus picture presentation was also manipulated within the RL paradigm. Explanations in terms of storage and retrieval advantages for high-imagery items and pictures have already been advanced within the free-recall paradigm (Paivio, Rogers, & Smythe, 1968). Because memory for past instances is influential in an RL reception paradigm, picture availability and accessibility should produce faster RL rates than words.

Method.—The design of the present experiment can be described as a $2 \times 4 \times 2 \times 2$ factorial with group (geometric vs. transportation stimuli), rule difficulty (CJ, DJ, CD, BCD), stimuli presentation (picture vs. word), and RV congruency (congruent vs. incongruent) as the four independent variables. The *Ss* were 128 volunteers from an introductory psychology course at the University of Texas at Austin, with random assignment of four *Ss* to each experimental condition. All *Ss* were run individually and were naive with respect to verbal learning and concept learning experiments.

---

[1] The author wishes to express his appreciation to Lyle E. Bourne, Jr., and to an anonymous reviewer for their invaluable comments on an earlier draft of the present paper.

[2] Requests for reprints should be sent to A. Keith Barton, now at the Department of Psychology, Austin State School, Box 1269, Austin, Texas 78767.

[3] The application of verbal learning variables to concept learning has been studied extensively for many years. Specifically, the introduction of "associative relatedness" may be traced to the earlier studies involving paired-associate paradigms, sense impressions, and dominance levels (e.g., Underwood & Richardson, 1956). Likewise, the pictorial manipulation parallels the same tradition as the Heidbreder (1946) "concrete objects." The present verbal learning application, however, is to rule learning, a subset of concept learning.

Four dimensions with three attributes or values per dimension were employed in all rule-learning problems with number (1, 2, 3), size (small, medium, large), and color (red, blue, green) common to all problems. In addition, shape (square, triangle, circle) and transportation object (firetruck, boat, airplane) were employed in the geometric shape (GS) and transportation category (TC) conditions, respectively. Color and shape (red–square) and color and transportation object (red–firetruck) were the RVs for Cond. GS Congruent (GS-C) and TC Congruent (TC-C) rule-learning problems, respectively. Likewise, green–triangle and green–boat were the RVs in the incongruent conditions (GS-I and TC-I).

It can be seen that the congruency–incongruency distinction is arbitrary in the GS condition while there is some basis for the distinction in the TC condition. It should be noted, however, that only one example per congruent and incongruent condition was used in the present study. Cond. GS is also a replication of typical stimulus populations used in RL experiments, while Cond. TC represents another dimension to allow for the possible influence of prior AV and imagery effects of the RVs.

Stimuli were presented to *Ss* on 5 × 9 white index cards following the reception paradigm. In the word (W) condition, four words were typed in all capitals in the middle of the index cards, always in the order of number, size, color, and shape or transportation object. A corresponding set of 81 stimulus cards was used in the picture (P) condition. As is typical procedure in RL experiments, all *Ss* were fully informed of the entire stimulus population to be used in the experiment. The names of the four dimensions with three values per dimension were typed on separate index cards and remained in view of *Ss* during the experiment. Likewise, the RVs for Cond. GS-C, GS-I, TC-C, TC-I were specifically mentioned to *Ss* and these values also remained in full view of *Ss* during the experiment.

In all eight conditions (P and W conditions for Groups GS-C, GS-I TC-C, and TC-I), stimulus cards were presented in the same order to equate positive–negative instance input within congruent and incongruent conditions—e.g., if "two large red firetrucks" was the tenth card in TC-C in both P and W conditions, then "two large red squares" was also the tenth card in GS-C. Likewise, "two large green boats" and "two large green triangles" were also the tenth cards in TC-I and GS-I, respectively.

Each card presented corresponded to 1 trial, and a criterion of 16 consecutive correct responses or 100 trials was employed. The *S* responded YES or NO to each card as to whether or not *S* thought the card was a positive or negative instance of the concept based on the rule *S* was trying to learn. Feedback from *E* (right–wrong) was provided to *S* after each trial. The *Ss* had as much time to respond to a given card with an average response latency of approximately 7 sec. in presolution phase. If more than 81 trials (cards) were required for solution, the first 19 cards were administered again in the same order. Each *S* was presented with only one problem.

Results and discussion.—Trial of last error was used to assess RL rates for all groups. As expected, rule difficulty provided the largest source of variance, $F\ (3, 96) = 51.15$, $p < .001$. However, the actual ordering was CJ < DJ < BCD < CD, with means of 9.06, 22.80, 60.90, and 75.50, respectively; a finding at some variance from Bourne (1970). Subsequent Newman-Keuls tests confirmed that CJ < DJ and that BCD < CD, $p < .01$. Employing an easy–difficult rule dichotomy, a Scheffé test confirmed that CJ and DJ rules were easier than CD and BCD rules, $p < .01$. Pictures were also found to be easier than words, $F\ (1, 96) = 4.01$, $p < .05$. The main effect of P versus W is localized at the BCD rule, although the Rule Difficulty × P versus W interaction does not approach significance, $.05 < p < .10$. The only remaining source of variation was the four-way interaction including all four variables, $F\ (3, 96) = 3.06$, $p < .05$.

The results of the present report replicate earlier findings (Bourne, 1970; Haygood & Bourne, 1965) in that learning rates are retarded by increasing rule complexity. However, the CD rule was found the most difficult in the present study, a finding at variance with the above studies which found the BCD rule the most difficult. The DJ and BCD means are in close agreement with Bourne's Problem 1 means, although the CJ and CD rules are discrepant.

The CD and BCD rule reversal observed in the present study is an interesting question in itself and deserves further note. The present results establishing CD rule learning as the most difficult are by no means unique (Giambra, 1970; Taplin, 1971). However, recent evidence (Sawyer & Johnson, 1971) has discounted previous claims (Giambra, 1970; Taplin, 1971) that the CD-BCD reversal is task dependent (selection vs. reception mode) or initial instance dependent (positive vs. negative instance).

This report makes no major attempt to resolve the CD-BCD rule controversy, but an empirical finding is supportive of Sawyer and Johnson's (1971) earlier critique of Giambra (1970). The GS-C and GS-I conditions of the present experiment represented initial positive and negative instances, respectively, for CD rules. These two cells are typical of previous stimulus populations employed in rule-learning research, and the positive–negative initial instance is not confounded with the congruency–incongruency manipulation (as is the case in Groups TC-C and TC-I). As has been suggested before (Bourne, 1970), negative (TF) instances are extremely critical to CD rule ascertainment; however, the initial negative instance of the GS-I cell for the CD rule took a mean of 20 trials longer than the initial positive instance (FF) of the GS-C cell. A Newman-Keuls test did not confirm this finding however, $p > .05$, but the point is that initial negative instances most definitely did not improve CD performance.

An apparent explanation for the above CD and BCD rule difficulty reversal may reside in the stimulus presentation procedure used in the present study, in that the natural stimulus population was employed. Thus, all 81 different stimuli were presented before any 1 stimulus was repeated. The above manipulation imparted a 1:2:2:4 ratio for TT, TF, and FT, and FF instances, respectively, throughout the 81 trials for CD rule-learning problems. Perhaps, a constrained stimulus presentation which equalizes the number of stimuli from each truth-table class may have eliminated the CD-BCD reversal in that FF instances are decreased from 44% to 25% (L. E. Bourne, personal communication, 1971). Another crucial variable for the CD-BCD rule reversal may reside in the nonrandom sequencing of TT, TF, FT, and FF instances.

Apart from the CD-BCD rule controversy, the present experiment is also concerned with imagery and prior AV of the RVs. The imagery effect was evident from the nine-trial picture advantage over words. One explanation may be the reduced memory load due to the availability and accessibility of past instances with pictures. Contrary to previous expectations, prior AV of RVs seems of little significance insofar as GS and TC groups did not differ nor did they interact with rule difficulty, both $Fs < 1$.

In conclusion, present results support one of the two previously mentioned hypotheses—that pictures produce faster RL learning rates than words. The contrary results concerning RV "relatedness" may be due to the earlier suggestion (L. E. Bourne, personal communication, 1971) that such stimuli may have induced unit processing of the RVs, thereby impeding the acquisition of a truth-table coding strategy. In any case, a major intent of the present study was to focus more attention on other variables in addition to those currently being studied in rule-learning problems.

REFERENCES

BOURNE, L. E., JR. Knowing and using concepts. *Psychological Review*, 1970, **77**, 546–556.

GIAMBRA, L. M. Conditional and biconditional rule difficulty with attribute identification, rule learning, and complete learning task. *Journal of Experimental Psychology*, 1970, **86**, 250–254.

HAYGOOD, R. C., & BOURNE, L. E., JR. Attribute- and rule-learning aspects of conceptual behavior. *Psychological Review*, 1965, **72**, 175–195.

HEIDBREDER, E. The attainment of concepts: I. Terminology and methodology. *Journal of Genetic Psychology*, 1946, **35**, 173–189.

PAIVIO, A., ROGERS, T. B., & SMYTHE, P. C. Why are pictures easier to recall than words? *Psychonomic Science*, 1968, **11**, 137–138.

SAWYER, C. R., & JOHNSON, P. J. Conditional and biconditional rule difficulty under selection and reception conditions. *Journal of Experimental Psychology*, 1971, **89**, 424–426.

TAPLIN, J. E. Effect of initial instance on attribute identification of concepts using a selection procedure. *Journal of Experimental Psychology*, 1971, **88**, 177–181.

UNDERWOOD, B. J., & RICHARDSON, J. Verbal concept learning as a function of instruction and dominance level. *Journal of Experimental Psychology*, 1956, **51**, 229–238.

(Received December 23, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 211–213

INFORMATION PURCHASE STRATEGIES OF GAMBLERS AND STUDENTS[1]

KURT J. SNAPPER,[2] WARD EDWARDS, AND CAMERON R. PETERSON
University of Michigan

Two experiments were conducted to determine, first, whether *S*s' information purchase strategies were invariant under additive transformations of the payoff matrix and, second, whether the effect of additive transformations interacted with experimental setting. Two different experimental settings were used: Exp. I used casino players in Las Vegas, Nevada, and Exp. II used students at the University of Michigan. The additive transformations had a significant effect only in Exp. I.

The experimental literature, bearing on static models of risky decision making, exhibits a conflict between theoretical purism and realism (Edwards, 1961). The purist wants *S* to remain in a fixed financial state, experiencing no actual wins or losses, or has him play for trivial stakes. The realist, on the other hand, feels that *S*s respond to such unrealistic conditions by inventing strategies that permit rapid, thoughtless completion of the task (Slovic, Lichtenstein, & Edwards, 1965). Such artifactual strategies can usually be avoided only by ensuring that decisions have immediate, substantial consequences; this is often impossible in a university setting, in which stakes are usually small and *S*s (correctly) assume that they receive some minimal payments, regardless of *E*'s instructions to the contrary (e.g., Argyris, 1968).

A series of experiments at the Four Queens Casino in Las Vegas, Nevada, attacked some of these difficulties associated with university experiments. This paper reports the results of one of these experiments, involving the play of an information purchase game. The *S*s were casino customers who volunteered to play.

A second experiment was conducted at the University of Michigan in order to investigate whether the university setting would affect the results; this second experiment was an attempt at replicating the Las Vegas study, although some changes were necessary.

In both experiments, *S*s played a game requiring a choice between two alternative predictions. If *S* predicted correctly, he was paid a specified amount, but if he predicted incorrectly he was required to pay a specified amount. Prior to making a terminal prediction, *S* could sequentially purchase information that should reduce uncertainty about which prediction is correct. Edwards (1965) has shown that the optimal strategy is to select an odds criterion, to purchase information until the odds that one prediction rather than the other will be correct equals the criterion, and then to make the favored prediction. The optimal odds criterion, Ω^*, depends upon the difference between the amounts that can be won or lost, the cost per datum, and the diagnostic value of data. Because V, the difference between amounts that can be won or lost, is not uniquely determined, many sets of payoffs lead to the same Ω^*. In particular, the addition of a constant amount to both the amount to be won and the amount to be lost does not alter V, and therefore does not change the optimal strategy; V was constant in both experiments, but component payoffs were varied by addition of a constant to both the amount of a win and a loss.

Experiment I.—A modified bingo basket contained, with equal probability, six red and four blue marbles, or four red and six blue marbles. The *S*s played with poker chips worth $.10 each. The *S*s could observe marbles sampled sequentially with replacement from the bingo basket at a cost of one chip per observation; *S* could continue to purchase observations until deciding to make a prediction about the predominant color in the basket.

Two conditions, I_A and I_B, were used. The *S*s in Cond. I_A encountered only Payoff A, for which the amounts to be won and lost were, respectively, 25 and -25 chips. The *S*s in Cond. I_B encountered only Payoff B, for which the amounts to be won and lost were, respectively, 20 and -30 chips. In both conditions, the optimal strategy is to purchase observations until $S - F$, the difference between the number of marbles in the sample of each color, equals two. When an $S - F$ of two is reached *S* should terminate sampling and predict that the majority marble color in the basket is the same as the color of the majority of the marbles in the sample. The expected value for the optimal strategy is 5 and 0 chips for Cond. I_A and I_B, respectively. A within-*S* design was not feasible in Exp. I. The game was held in a public area where prospective *S*s could observe which conditions were being used. Experience with a pilot study suggested that *S*s would not play for Payoff B when aware that by "passing" they could play for more attractive payoffs.

At the start of each trial, *E*, a professional dealer, drew a card at random from a deck of standard playing cards and, depending on the color of the card, pressed one of two concealed keys connected

---

[1] The research reported here was sponsored by the Wood Kalb Foundation. Space and facilities for the research were made available by the Four Queens Hotel and Casino in Las Vegas, Nevada. We are deeply grateful to both. We are also grateful to Barbara E. Goodman and Gregory Fischer for their valuable help.

[2] Requests for reprints should be sent to Kurt J. Snapper, University of Michigan, HSRI Building, 2901 Baxter Road, Ann Arbor, Michigan 48105.

to a PDP-7 computer. The computer was connected to a marble-counting apparatus which counted the appropriate marbles into the basket, depending upon which key the dealer pressed. The bingo basket was specially modified so that randomly selected marbles could be singly viewed apart from the other marbles. The dealer then gave the player the option of sequentially purchasing observations. When the player made his terminal prediction he either paid or was paid the appropriate number of chips. The dealer relayed all information about play of the game to the computer; the computer stored this information for later analysis at the University of Michigan.

The *S*s were customers of the Four Queens Casino who volunteered to play the game; they were free to quit at any time. All *S*s played the game individually, with chips purchased from the casino; payment was the amount *S* won or lost during the game. Twelve and 10 *S*s, respectively, served in Cond. I_A and I_B.

The mean values of $S - F$, at which *S*s stopped buying observations, were 1.46 and 2.61 for conditions I_A and I_B, respectively. The difference between group means was significant by *t* test at the .01 level. The corresponding mean number of observations per trial were 3.40 and 4.42, respectively.

The median number of trials played was only seven. Five *S*s from each condition quit the game as losers. Verbal reports to the dealer indicated that most *S*s purchased an average of about 100 chips, and quit the game if they lost this stake.

Experiment II.—Experiment II was conducted at the University of Michigan, with college students as *S*s. Three conditions were used. All conditions consisted of two blocks of seven trials each; blocks of seven were chosen since that was the median number of trials completed by Exp. I *S*s. In Cond. II_{AB}, the first block involved only Payoff A and the second block involved only Payoff B; this order was reversed in Cond. II_{BA}. In Cond. II_M, Payoffs A and B were presented on alternating trials, with the order of initial presentation counterbalanced. The first blocks from Cond. II_{AB} and II_{BA} are analogous to Cond. I_A and I_B from Exp. I.

The apparatus from Exp. I was not available for Exp. II; in addition some procedural and design changes were made. At the start of the session, each *S* was given a 100-chip stake and was informed that if he lost his stake he could not continue and would receive no payment; this procedure was used because many Exp. I *S*s apparently staked themselves a similar amount and quit when the stake was lost. The *S*s were also informed that any winnings above the amount of the stake would be their payment for participation; unknown to *S*s, they were paid $1.00 minimum. Each chip was worth $.05 rather than $.10 in order to reduce the variance in *S*s' winnings to an acceptable level. Two urns, with marble compositions identical to those used in Exp. I, were displayed and explained to *S*. The *E* then spun a roulette wheel to select, with equal probability, one urn; *S* was informed about all contingencies governing urn selection. The urns and roulette wheel were concealed from *S* behind a translucent screen. The *E* then randomly sampled marbles until *S* made a terminal prediction. Both *S* and *E* independently recorded the color of each observation. On the earliest trial the second payoff scheme was presented, *S* was informed about the new payoff contingencies and was reminded on subsequent trials. No *S* was ever given any indication that two payoff schemes would be used until the second scheme was presented on the earliest appropriate trial.

Thirty male students from the University of Michigan served as *S*s and 10 were randomly assigned to each condition. Payment ranged from the $1.00 minimum to $8.75.

The mean values of $S - F$, calculated from means for individual *S*s, from the first blocks of Cond. II_{AB} and II_{BA} were, respectively, 1.99 and 1.83. For comparison, the means from Cond. I_A and I_B, for only the first seven trials (or, if less than seven, the number of trials actually completed) were 1.37 and 2.66. The effects of Payoffs A and B in Exp. II were opposite in direction to the effect observed in Exp. I. An analysis of variance, based on all the Exp. I data and the first blocks from Cond. II_{AB} and II_{BA}, revealed a significant interaction, $F(1, 38) = 9.68$, $p < .01$. Mean values of $S - F$ from the second blocks of conditions II_{AB} and II_{BA} were, respectively, 2.23 and 1.91; mean values associated with Payoffs A and B in Cond. II_M were 1.78 and 1.83, respectively. None of the between-block differences was significant. The overall mean number of observations per trial was 4.11 in Exp. II; differences were not significant between any blocks.

Only one *S*, from Cond. II_{AB}, lost his stake, after completing 13 of 14 trials.

Discussion.—Unfortunately, several factors differed between Exp. I and II. While any of these could conceivably account for the observed interaction, the most plausible explanation is that casino *S*s are accustomed to games in which payoffs are closely linked to the odds of winning. In a fair game, if the odds are 4:1 against winning, a fair bet would have a payoff ratio of four times the stake. (Of course, the stake is returned to the player if he wins.) While few games in Las Vegas are fair in this sense, almost all are almost fair, and knowledge of this principle is widespread among casino customers.

The casino players' behavior in Exp. I could reflect this principle. As it happens, however, the ratio is not relevant to optimal information purchase strategies; recall that it is the difference, not the ratio, of the amounts to be won and lost that is relevant. The payoff ratios were 1:1 and 2:3 in Cond. I_A and I_B, respectively; but the odds of winning are controlled by the purchase of information. It seems quite natural that *S*s, playing an unfamiliar game, might demand relatively higher odds—and therefore purchase more information—in Cond. I_B than in Cond. I_A. University of Michigan students, on the other hand, were little affected by the payoff differences. The student may have understood the formal properties of the game better; more likely, they were too ignorant of conventional gambling lore to be misled by it.

REFERENCES

ARGYRIS, C. Some unintended consequences of rigorous research. *Psychological Bulletin*, 1968, **70**, 185–197.

EDWARDS, W. Behavioral decision theory. *Annual Review of Psychology*, 1961, **12**, 473–498.

EDWARDS, W. Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, 1965, **2**, 312–329.

SLOVIC, P., LICHTENSTEIN, S. C., & EDWARDS, W. Boredom induced changes in preferences among bets. *American Journal of Psychology*, 1965, **78**, 208–217.

(Received January 5, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 213–215

NOUNS, ADJECTIVES, AND SEMANTIC MEMORY[1]

ELIZABETH F. LOFTUS[2]
New School for Social Research

Two hundred *S*s were presented with 20 nouns and 20 adjectives. If the stimulus word was a noun category, *S*s were told to write any item that belonged in that category (e.g., SEAFOOD—OYSTER, CLAM, FISH, etc.). For adjectives, *S*s were told to write any item which had the inherent quality of that adjective (e.g., HARD—BRICK, ROCK, WOOD, etc.). The mean number of responses given to noun stimuli was greater than the mean number of responses given to adjective stimuli. The number of responses given was found to be strongly related to the frequency of the stimulus for nouns, but not for adjectives. The data support the notion that the semantic memory store is organized primarily into noun categories.

In an earlier study, Freedman and Loftus (1971) presented *S*s with a noun category plus a restricting letter or adjective and asked *S*s to produce an instance of the category which satisfied the restriction. A general model of the retrieval process postulates at least two major steps: (*a*) entering the category, and (*b*) finding an appropriate member of that category. In this model, the assumption is made that the memory store is organized primarily into noun categories. This assumption is not meant to deny the existence of groups of items having qualities in common (e.g., small, green), but the major organization is thought to be in terms of noun groupings.

The assumption that memory is organized into nouns has several important implications. One observable consequence concerns the number of words that *S* can produce in a given amount of time under various conditions. If we present a noun category and give *S*, say, 1 min. to write down any items that belong in the category, he should be able to produce more items than if we present him with an adjective and ask him to write any items which have the inherent quality of that adjective. According to the model, when *S* is presented with a noun category, he enters the appropriate category in his semantic memory and begins to name members of the category that are stored there. When he sees an adjective, however, he must still enter the memory store at some noun category, whereupon he can name members of that category which have the inherent quality of the presented adjective. When he has named as many members as he can, he must shift to another category and begin to name items from it. Thus, when presented with an adjective, fewer items should be produced for two reasons: (*a*) *S* must decide which category to enter first, and this decision takes time; and (*b*) *S* must shift from one category to another whenever he needs to, and shifting takes time.

The present experiment was designed to determine whether the number of responses given to a noun is different from the number of responses given to an adjective.

Method.—The *S*s were 200 students at Stanford University and the New School for Social Research. A group-testing procedure involving a booklet technique was used, with groups ranging in size from 3 to 48 *S*s. The *S*s were given notebooks with a stimulus word written on the top of each page. These words were the nouns and adjectives shown in Table 1. Twenty nouns and 20 adjectives were selected with two criteria in mind. First, they were chosen so as to include a wide range of stimulus frequencies (Kučera & Francis, 1967). The nouns ranged in frequency from a low of 37 (profession) to a high of 808 (state). The adjectives ranged in frequency from a low of 37 (enormous) to a high of 755 (long). The mean frequency for nouns was 176.6 while for adjectives it was 213.5. Second, nouns and adjectives that had appeared in previous studies and norms were selected, so as to provide

---

[1] This research was supported by National Institute of Mental Health Grant MH-20280.

[2] Requests for reprints should be sent to Elizabeth F. Loftus, New School for Social Research, 66 West 12th St., New York, N. Y. 10011.

TABLE 1

Frequency in English and Median Number of Responses Given to Twenty Nouns and Twenty Adjectives

Nouns			Adjectives		
Noun	Frequency	Median response	Adjective	Frequency	Median response
activity	116	13	enormous	37	8
animal	68	16	fast	78	9
artist	57	7	foreign	158	9
author	46	8	French	139	10
car	274	15	green	116	9
city	393	15.5	heavy	110	9
cloth	43	9	high	497	9
color	141	15	hot	130	11
country	324	16.5	large	361	9
food	147	15	living	194	9
furniture	39	11	long	755	8
metal	61	10	male	37	9
month	130	12	modern	198	9
president	382	11	primitive	38	8.5
profession	37	11	religious	165	9.5
season	105	5	short	212	7
snake	44	8	small	542	10
state	808	14.5	white	365	10
weapon	42	12	winter	83	11
word	274	16	yellow	55	9
$\bar{X}$	176.6	12.03	$\bar{X}$	213.5	9.15

some continuity with previous research.[3] A new randomization of these stimuli was used for each *S*.

The *Ss* were told that each stimulus would be either a noun or an adjective. If the stimulus word was a noun category, *S* was to write down as many items as he could belonging to that category. For example, *Ss* who were presented with the stimulus SEAFOOD might say OYSTER, CLAM, or SHRIMP, among other possibilities. If the stimulus was an adjective, *S* was to write down as many items as he could that had the inherent quality of the adjective. For example, if the stimulus was HARD, *S* might say BRICK, ROCK, or WOOD. The procedure was as follows: *E* said "begin"; *S* turned the page of his notebook and was given 1 min. to read the name of the stimulus and write down as many items as he could, in whatever order they happened to occur to him. A buzzer ended each 1-min. trial.

Results and discussion.—In the following analyses, all separate responses to the 40 items were counted. This means that in the case of nouns we counted all responses even when they did not seem to belong to the noun category, and in the case of adjectives we counted all responses even though some did not have the inherent quality of the adjective.[4] Table 1 presents the median number of responses given to each stimulus. Means were obtained by averaging medians for the two types of stimuli. The mean number of responses given to noun stimuli (12.03) was significantly greater than the mean number of responses given to adjective stimuli (9.15), $t(38) = 2.18$, $p < .05$. This result is consistent with a model that describes memory as being organized primarily into noun categories. More detailed analyses of the response protocols support this general view. When the stimulus was a noun, almost all of the *Ss'* responses were instances of that noun category. There was some tendency to group together category members which have qualities in common. For example, SWEDEN, NORWAY, and FINLAND (all Scandinavian countries) often occurred together in the response protocols for COUNTRY. It appears that within many categories, a variety of subsets often exist—subsets whose members are highly associated perhaps as a result of the fact that they have these qualities in common. These subsets seem to be more idiosyncratic than the main, noun categories, but are somewhat consistent across individuals. In contrast, when the stimulus was an adjective, *Ss* tended to name a number of items from one noun category, shift to another category and name a number of items from that category, shift again, and so on. An actual sample protocol for the stimulus SMALL best illustrates this point: FLY, BEE, ANT, EYE, EAR, TOE, FINGER, GRAPE, CHERRY, STRAWBERRY. The *S* named three insects, four body parts, and three fruits. Apparently the reason *Ss* name fewer items in the case of adjectives is that they use up time shifting from one category to another. In the case of nouns, any "shifting" which may occur is confined within the main noun category itself.

A second result concerns the effect of stimulus frequency on the number of responses *S* produces in 1 min. The frequency in English, according to Kučera and Francis (1967), for each stimulus is given in Table 1. There was a strong correlation between frequency and number of responses for noun stimuli, $r = .47$, $p < .05$. This result is consistent with an earlier finding that the speed of producing a single category member was faster when high rather than low frequency category names were used, and the effect of frequency persisted even when the time taken to read and identify the category names was removed (Loftus & Freedman, 1972). It is tempting to explain these results in terms of association strengths between the category name and the instance produced. The notion that some category names are more strongly associated to an instance than are other names has already received some experimental support in the work of Loftus and Scheff (1971). For example, the instance BUTTERFLY is more strongly related to high frequency superordinate INSECT than it is to the lower frequency superordinate BUG.

For adjective stimuli, there was no relationship between word frequency and number of responses, $r = .09$, $p > .10$. According to the general model discussed in the introduction, the first step of the retrieval process involves entering the noun category. Only after *S* has entered a noun category does he find clusters of items that have qualities in common (e.g., Scandinavian countries). In other words, after completing the first step of the process

---

[3] We did not realize until after the experiment was completed that two of the adjectives, MALE and WINTER, might have confused *Ss*. Both of these words are homographs; they could have been interpreted either as nouns or adjectives. In any event, the reader has Table 1 available to him, and may see for himself what happened in these two cases. There was no indication that this was a problem for any of the *Ss*.

[4] For some of the adjectives, many *Ss* responded as if they had interpreted the instructions to read "anything which *sometimes* does have the quality of the adjective." For example, the words BUS, FLOWER, and LIGHT occurred in response to YELLOW.

S may find a cluster of instances to which a given adjective applies. It is puzzling that adjective frequency has no effect, but for the moment we have no explanation.

REFERENCES

FREEDMAN, J. L., & LOFTUS, E. F. Retrieval of words from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 107–115.

KUČERA, H., & FRANCIS, W. N. *Computational analysis of present-day American English.* Providence: Brown University Press, 1967.

LOFTUS, E. F., & FREEDMAN, J. L. Effect of category-name frequency on the speed of naming an instance of the category. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 343–347.

LOFTUS, E. F., & SCHEFF, R. W. Categorization norms for fifty representative instances. *Journal of Experimental Psychology*, 1971, **91**, 355–364. (Monograph Supplement)

(Received January 29, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 215–218

FAILURES OF CONTINGENCY AND COGNITIVE FACTORS TO AFFECT LONG-INTERVAL DIFFERENTIAL PAVLOVIAN AUTONOMIC CONDITIONING[1]

KARL SCHIFFMAN AND JOHN J. FUREDY[2]

University of Toronto, Ontario, Canada

According to the "contingency" prediction, in a two-group differential conditioning design, discrimination between CS+ and an explicitly unpaired CS− should exceed discrimination between CS+ and a "truly random" CS−, a stimulus which is uncorrelated with UCS occurrence. This did not occur in 40 *S*s run under an 8-sec. CS–UCS delay conditioning paradigm with GSR and digital volume-pulse change (VPC) as dependent autonomic variables. Moreover, cognitive subjective-contingency (SC) measures indicated that *S*s were aware of the contingency difference between the two types of CS− stimuli. Finally, despite the fact that the long CS–UCS interval was similar to those employed by investigators who have recently reported that knowledge of CS–UCS relationships is important for differential autonomic conditioning, the results, contrary to this "cognitive" position, did not yield any correlation between autonomic (GSR and VPC) and cognitive (SC) differentiation (CS+ vs. CS− performance).

It has been indicated previously (Furedy & Schiffman, 1971) that the operation of contingency factors (Rescorla, 1967) in differential Pavlovian autonomic conditioning implies the *contingency prediction:* that if reliable CS+:CS− discrimination is obtained in a two-groups differential conditioning design, discrimination between CS+ and an explicitly unpaired CS− (euCS−) will exceed discrimination between CS+ and a "truly random" CS− (trCS−), where trCS− is not negatively correlated, but merely uncorrelated, with the occurrence of the unconditional stimulus (UCS).[3] The *cognitive prediction*, which follows from the view that cognitive factors affect differential autonomic conditioning (e.g., Wilson, 1968), is best stated as the assertion that: "verbalised knowledge of CS–UCS relations is related to the extent of differential conditioning [Baer & Fuhrer, 1970, p. 178]". In a recent test of the two predictions, neither was confirmed with the galvanic skin response (GSR) and digital volume-pulse change (VPC) as dependent autonomic variables (Furedy & Schiffman, 1971).

---

[1] This study was based on a thesis submitted in partial fulfillment of the requirements for the master's degree at the University of Toronto by the senior author under the supervision of the junior author. This research was briefly reported at the meeting of the Psychonomic Society, St. Louis, November 1971, and was supported by grants from the National Research Council (APA 222) and from Canada Council (S70-0710).

[2] Requests for reprints should be sent to John J. Furedy, Department of Psychology, University of Toronto, Toronto 181, Canada.

[3] This prediction is clearly derivable from Rescorla's stated application of the contingency position to differential Pavlovian conditioning (Rescorla, 1967, p. 74), but a later statement that "the contingency approach *only* predicts this difference (between euCS− and trCS−) when the CS is tested in the presence of some other excitatory stimulus [Rescorla, 1967, p. 78] may, at first sight, appear to negate the present and previous (Furedy & Schiffman, 1971) methods of testing the contingency position's applicability to autonomic conditioning. Specifically, on this view, it would appear that a compound-stimulus design (with jointly presented CS+ and CS− trials) would be required to test the contingency position. This view is probably correct when applied to the nonautonomic responses which have concerned Rescorla and, with respect to which the CS, before conditioning, is "neutral" in the sense of producing no measurable responding. Thus, the inhibitory properties on shuttling behavior that may be acquired by a tone being negatively correlated (i.e., an euCS−) with shock would not be registered by simply measuring shuttling to the tone alone, since lower-than-no-shuttling behavior cannot be readily observed. For autonomic responses, however, it is well known that there are no such initially "neutral" CSs, since all stimuli elicit some measurable responding. Hence, in Rescorla's terms, even when an euCS− alone is presented, there is present "some other excitatory stimulus" in the sense that the euCS−, qua stimulus, does elicit some level of measurable responding. Accordingly, a compound-stimulus design is not necessary to test the contingency prediction in *autonomic* Pavlovian conditioning; nor is such a design desirable since the data on this more complex form of autonomic conditioning are both less clear and less plentiful than data from studies where the various CSs are presented alone.

That test, however, employed a CS–UCS interval of .75 sec. in order to ensure reliable CS+:CS− discrimination (Furedy, 1970). Because such short CS–UCS interval arrangements for autonomic conditioning require the use of CS+ alone test trials to evaluate CS+:CS− discrimination, some investigators believe, as detailed elsewhere (Furedy, 1968, pp. 70–71), that any obtained CS+:CS− difference can be accounted for solely in terms of orienting-reaction (OR) reinstatement to change from paired CS+UCS trials to omitted-UCS CS+ test trials. Accordingly, in the present test of the contingency prediction, a CS–UCS interval of 8 sec. was used to allow CS+:CS− discrimination to be measured independently of any OR effects due to change from CS + UCS to CS+ alone trials. Moreover, the use of a longer CS–UCS interval also served to eliminate a potentially relevant difference between previous confirmations of the cognitive prediction (e.g., Baer & Fuhrer, 1970) which have generally employed CS–UCS intervals of about 8 sec., and Furedy and Schiffman's (1971) disconfirmation of the prediction, a disconfirmation based on the shorter .75-sec. CS–UCS interval. Finally, another potentially relevant aspect of the confirmatory studies is that the CS+ and CS− in these experiments have typically been in the same modality (usually two tones of differing frequency), whereas in all the autonomic differential conditioning studies run in the Toronto laboratory, in order to maximize CS+:CS− discrimination, the modality of the two CSs has been varied with tone and light as the counterbalanced CS+ and CS−. Accordingly, in the present experiment, the CS modality factor was varied between *S*s to see whether the cognitive prediction would be more clearly confirmed with the unimodally than with the multimodally varied CSs.

Method.—The *S*s were 40 undergraduates from the University of Toronto, who participated either to fulfill a course requirement or for $1.50.

The apparatus, including the visual CS, was the same as that used by Furedy and Schiffman (1971), except that two auditory CSs were employed. Both CSs were 79-db. SPL delivered through a speaker above *S*s head over a continuous 70-db. background white noise, but the high and low tones were 3,500 and 800 Hz., respectively.

The procedure employed 8-sec. CSs which, on CS + UCS trials, were followed immediately by the .3-sec. UCS (2.0-ma. shock) in a delay-conditioning paradigm. For half of the *S*s (multimodal-CSs condition), the 3,500-Hz. tone and the light served as CSs, with the nature of CS (tone vs. light) being counterbalanced. For the remaining 20 *S*s (unimodal-CSs condition), the two tones served as CSs, with the nature of CS (3,500 Hz. vs. 800 hz.) being counterbalanced. Before conditioning, all *S*s received a series of separate CS and UCS trials comprising two each of the CS+ and CS− trials interspersed with three UCS-alone trials; the intervals between these preliminary trials varied randomly between 30, 40, and 50 sec. Then, as in Furedy and Schiffman (1971), the conditioning trials series for the 20 *S*s in the euCS− group comprised 12 pairings of CS+ with UCS, 15 euCS− trials, and 3 CS+ alone trials, presented, respectively, after 0, 6, and 12 CS + UCS trials. From *S*'s viewpoint, the nature of the trials was unsystematically ordered, and the intertrial intervals varied randomly between 30, 40, and 50 sec. For the remaining 20 *S*s in the trCS− group, the conditioning-trials series was the same as for the euCS− group, except that the 15 euCS− trials were replaced by 15 trCS− trials. These trCS− trials, as in Furedy and Schiffman, were randomly allocated over the total conditioning session, so that trCS− occurrence was independent of, or uncorrelated with, UCS occurrence. In the present case, the allocation of the 15 CS− trials resulted in 2, 3, 5, and 5 occasions, respectively, on which a UCS followed a trCS− within 0–20, 20–40, 40–60, and 60–120+ sec., with the last trCS− occurring after the last UCS. This contrasts with the CS+, which was followed by a UCS after 8 sec. on 12 of 15 occasions, and with the euCS−, which was never followed by a UCS within 29 sec. The complete experimental design involved the orthogonal variation of the following between-*S*s factors: modality (multimodal vs. unimodal CS variation), nature of CS (light or 800-Hz. tone vs. 3,500-Hz tone), and type of CS− (euCS− vs. trCS−). These factors were counterbalanced with respect to the within-*S*s discrimination (CS+ vs. CS−) factor. Finally, in addition to the autonomic GSR and VPC, subjective contingency (SC) measures were obtained for CS+ and CS− at the end of the experiment, as detailed previously (Furedy and Schiffman, 1971, p. 162). Briefly, SC represents *S*'s estimate of the degree of correlation (positive or negative) between CS and UCS occurrence.

Results.—With the long CS–UCS interval used here, multiple autonomic responding can occur (e.g., Prokasy & Ebel, 1967). The latency criteria used for such responses were similar to the criteria employed in an earlier GSR and VPC conditioning study in which a 7.5-sec. CS–UCS interval was employed for one group of 80 *S*s (Furedy, 1970). Specifically, first-, second-, and third-interval GSRs were defined as drops of resistance occurring, respectively, 1–5, 5–9, and 9–13 sec. following CS onset; first- and second-interval VPCs were defined as vasoconstrictions occurring 2–8 and 10–16 sec. after CS onset. The micromho and percent-VPC transformations were used, respectively, for all GSRs and VPCs, as in Furedy and Schiffman (1971).

The contingency prediction was tested for autonomic responding in only those intervals where clear and reliable CS+:CS− discrimination was observed. This was so in the case of the first-interval GSR, in which preliminary-trials data yielded no significant difference between CS+ performance ($\bar{X} = 1.81$) and CS− performance ($\bar{X} = 1.74$), $F < 1$, but in which the conditioning-trials data yielded significantly higher CS+ performance ($\bar{X} = 2.32$) than CS− performance

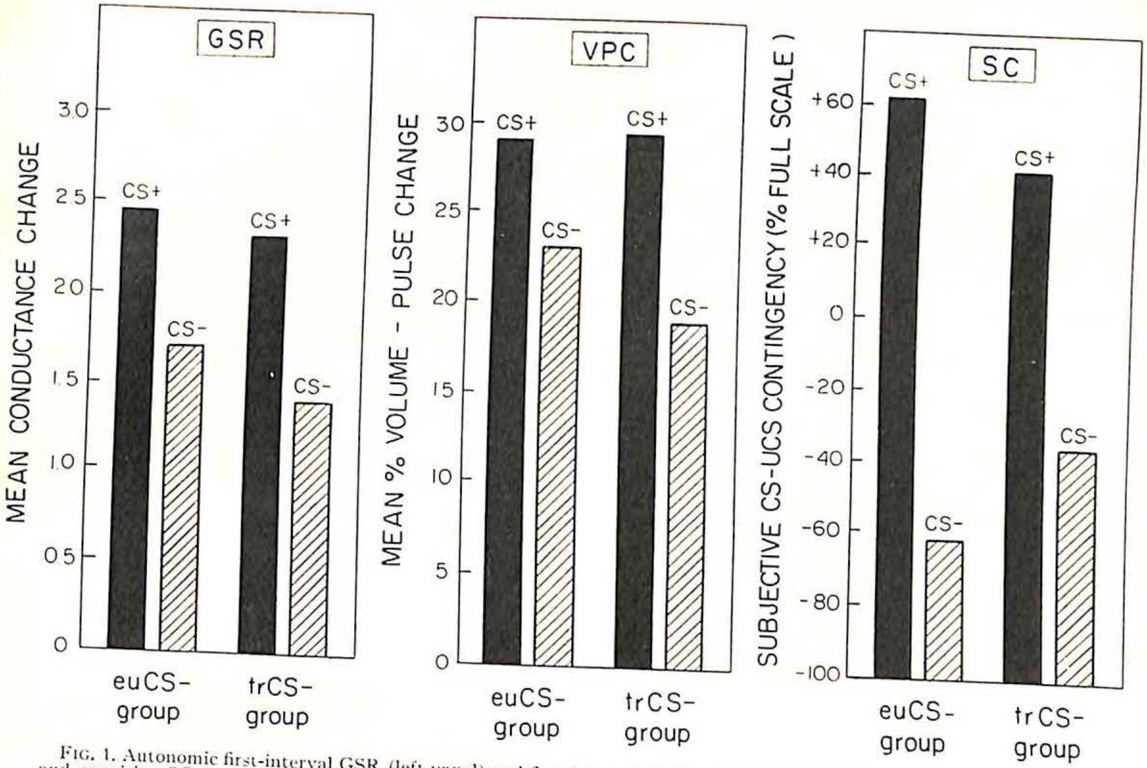


FIG. 1. Autonomic first-interval GSR (left panel) and first-interval VPC (center panel) measures during conditioning and cognitive SC (right panel) measures after conditioning on CS+ and CS− trials in the euCS− ($N = 20$) and trCS− ($N = 20$) groups.

($\bar{X} = 1.60$), $F\ (1, 32) = 41.82$, $p < .001$. Reliable CS+:CS− discrimination also emerged in the first-interval VPC, in which preliminary-trials data yielded no significant difference between CS+ performance ($\bar{X} = 30.17$) and CS− performance ($\bar{X} = 28.53$), $F < 1$, but in which conditioning-trials data yielded significantly higher CS+ performance ($\bar{X} = 29.25$) than CS− performance, ($\bar{X} = 21.22$), $F\ (1, 32) = 8.93$, $p < .01$. The modality factor (multimodal vs. unimodal CSs) did not interact with discrimination (CS+ vs. CS−) for any of the GSR or VPC responses, but there were interactions of modality with the nature-of-CS factor (light or 800-Hz. tone vs. 3,500-Hz. tone) during both the preliminary and conditioning stages. These interactions were in all cases due to the light and 800-Hz.-tone CSs producing equal responding, but the 3,500-Hz.-tone CSs producing a higher level of responding than either of the other two CSs.

The left (GSR) and center (VPC) panels of Fig. 1 show the data relevant to the contingency prediction: that CS+:CS− discrimination in the euCS− group exceed that in the trCS− group. Inspection of the panels indicates that even the trends were contrary to the contingency prediction for both autonomic measures, since for both measures the mean algebraic CS+:CS− difference in the trCS− group exceeded that in the euCS− group. These trends, however, did not approach significance for either measure, $t < 1$. A more generous test of the contingency prediction is to examine only late trials, since euCS− inhibition may be expected to develop over trials. Although such a test is not justified on statistical grounds, since there was no reliable Discrimination × Trials interaction during conditioning in either GSR or VPC, the data were, nevertheless, examined for mean performance over the last five conditioning trials (11–15). For GSR, these mean values were 2.45, 1.87, 2.37, and 1.35, respectively, for CS+ in the euCS− group, euCS−, CS+ in the trCS− group, and trCS−; for VPC, the corresponding values were 27.7, 22.5, 28.9, and 18.7. Inspection of these means indicates that, as in the case of the overall conditioning performance shown in Fig. 1, even the trends are contrary to the contingency prediction.

The SC results are shown in the right-hand panel of Fig. 1. As for the autonomic measures, clear cognitive discrimination of CS-UCS contingencies was obtained as regards the CS+:CS− difference, with CS+ being judged to be more positively correlated with the UCS than CS−, $t\ (39) = 12.1$, $p < .001$. However, in contrast to the autonomic trends, the CS+:CS− difference in the euCS− group did exceed the same difference in the trCS− group, $t(38) = 1.69$, $p < .05$, one-tailed. This SC outcome, unlike the autonomic outcomes, is consistent with the contingency prediction.

The cognitive prediction was tested for all autonomic measures by computing product-moment correlations between the extents of autonomic discriminations on the one hand and the extent of cognitive SC discrimination on the other hand. For all measures, extent of discrimination was defined as the algebraic difference between CS+ and CS− scores for each of the 40 Ss. The obtained correlations between SC and first-interval GSR, second-interval GSR, third-interval GSR, first-interval VPC, and second-interval VPC, respectively, were .10, −.12, .03, .07, and −.10. None of these correlations approached significance, and several were negative. Then, to see whether, as suggested in the introduction, the cognitive prediction would be more clearly confirmed with the unimodally varied CSs, the same five SC-autonomic correlations were computed for only those 20 Ss for whom the two tones served as the CSs. However, none of these correlations approached significance.

Discussion.—The first-interval GSR and VPC yielded reliable autonomic discrimination, but the contingency prediction was not confirmed. Indeed, as with the previous disconfirmation with the shorter CS–UCS interval (Furedy and Schiffman, 1971), even the trends in both autonomic measures (Fig. 1) were in the opposite direction to that expected on the basis of the contingency prediction. The failure to obtain reliable discrimination in the second-interval GSR is inconsistent with both the results of Prokasy and Ebel (1967) and of Furedy (1970); only the former investigators also obtained reliable third-interval GSR discrimination in contrast both to Baer and Fuhrer's (1970) and the present third-interval-GSR results; Furedy, but not Prokasy and Ebel, recorded second-interval VPC, and, and as in the present study, obtained no reliable discrimination in this autonomic measure. Taken together, these results suggest that post-first-interval responding in the GSR and VPC does not readily and universally yield reliable discrimination. Moreover, if the first-interval responses which did show reliable CS+:CS− discrimination here are conceived of only as ORs (e.g., Prokasy & Ebel, 1967, p. 255), then it follows that the contingency prediction was tested in the present instance only for conditional ORs. It bears emphasis, however, that the test could not have been confounded by any OR *reinstatement* effects resulting from the change from paired CS + UCS trials to omitted-UCS CS+ test trials, because, in contrast to the conditions of the previous study (Furedy & Schiffman, 1971), the long CS–UCS interval used here exceeded the measurement periods of both the first-interval autonomic responses in which reliable CS+:CS− discrimination was demonstrated.

The cognitive prediction was indirectly disconfirmed by the fact that the SC results, in contrast to the autonomic results, were consistent with the contingency prediction, which suggested that Ss were aware of the contingency differences considered important by the contingency position (Rescorla, 1967). However, the cognitive prediction was more directly disconfirmed by the consistent failures of any reliable positive correlations to emerge between the extent of cognitive and autonomic discrimination, even when these correlations were examined in the subset of Ss for whom the CSs were unimodally varied. The present disconfirmation of the cognitive prediction does not necessarily conflict with Baer and Fuhrer's (1970) finding that Ss differentiated in terms of accuracy of CS–UCS contingency; verbalizations were also differentiated in terms of differential autonomic responding. The apparent conflict between the present results and those of Baer and Fuhrer's (1970) is readily resolved by the possibility that the cognitive prediction holds only when Ss are split into accurate versus inaccurate verbalizers of CS–UCS contingencies. What the present disconfirmation of the cognitive prediction suggests is that such a dichotomous classification of the CS–UCS relational awareness dimension (Baer & Fuhrer, 1970) may obscure a breakdown of the cognitive-autonomic relationship over intermediate values of relational awareness, values which are tapped adequately only by the continuous SC measure used here and by Furedy and Schiffman (1971).

It is also worth noting that the generality of application of the present results is limited by the failure to obtain reliable second-interval GSR discrimination, especially since it has been suggested by some investigators (e.g., Fuhrer & Baer, 1969) that this response is more influenced by cognitive factors than is the first-interval GSR. On the other hand, as in Baer and Fuhrer's (1970) experiment, both first-interval GSR and first-interval VPC did show reliable autonomic discrimination. In conclusion, then, for those responses which did show reliable discrimination, both the cognitive and the contingency predictions were relatively unequivocally disconfirmed by the present results, which suggest that, as in the case of short-interval conditioning (Furedy & Schiffman, 1971), one should not assume that either contingency or cognitive factors necessarily affect long-interval differential Pavlovian autonomic conditioning.

REFERENCES

BAER, P. E., & FUHRER, M. J. Cognitive processes in the differential trace conditioning of electrodermal and vasomotor activity. *Journal of Experimental Psychology*, 1970, **84**, 176–178.

FUHRER, M. J., & BAER, P. E. Cognitive processes in differential GSR conditioning: Effects of a masking task. *American Journal of Psychology*, 1969, **82**, 168–180.

FUREDY, J. J. Human orienting reaction as a function of electrodermal versus plethysmographic response modes and single versus alternating stimulus series. *Journal of Experimental Psychology*, 1968, **77**, 70–78.

FUREDY, J. J. CS and UCS intervals and orders in human autonomic classical differential trace conditioning. *Canadian Journal of Psychology*, 1970, **24**, 417–426.

FUREDY, J. J., & SCHIFFMAN, K. Test of the propriety of the traditional discriminative control procedure in Pavlovian electrodermal and plethysmographic conditioning. *Journal of Experimental Psychology*, 1971, **91**, 161–164.

PROKASY, W. F., & EBEL, H. C. Three components of the classically conditioned GSR in human subjects. *Journal of Experimental Psychology*, 1967, **73**, 247–256.

RESCORLA, R. A. Pavlovian conditioning and its proper control procedures. *Psychological Review*, 1967, **74**, 71–80.

WILSON, G. D. Reversal of differential GSR conditioning by instructions. *Journal of Experimental Psychology*, 1968, **76**, 491–493.

(Received February 1, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 219–222

EFFECTS OF MODALITY AND SIMILARITY ON CONTEXT RECALL[1]

MICHELENE T. H. CHI[2] AND WILLIAM G. CHASE
Carnegie-Mellon University

In S. Sternberg's 1967 context-recall task, the search rate through short-term memory was a function of the sensory modality: If the memory set and test item were presented in the same mode (auditory–auditory or visual–visual), search rates were 48 msec. faster than for opposite-mode presentations.

In Sternberg's (1966) item-recognition task, Chase and Calfee (1969) found that when the memory set and test item were presented in different modalities (auditory vs. visual), search time was consistently slower than when the same mode was used. The slower search rates in the cross-modal case suggest that some of the original sensory information still remains in short-term memory even when verbal materials are encoded. Studies by Posner and his colleagues (e.g., Posner & Keele, 1967), using a same–different judgment, support this assumption. That is, responses to items presented in the same mode (within a few seconds of each other) are always faster than those in mixed modes.

Chase and Calfee (1969) further failed to find a consistent effect of acoustic confusability, although Conrad (1964) and many others have shown that acoustic similarity is a major source of interference in short-term memory tasks requiring recall of verbal materials. These effects of acoustic interference at recall occur regardless of whether the original material is presented visually or auditorily. Chase and Calfee, however, reported slower search rates through acoustically confusing material in only one of three item-recognition experiments.

The implication of Chase and Calfee's (1969) findings is that in the item-recognition task, the comparison does not always involve an acoustic-articulatory match but rather a sensory match is made if the modes are the same, and a higher level comparison is made if the modes are mixed. Chase and Calfee reconciled their findings with Conrad's (1964) by suggesting that perhaps the item-recognition paradigm does not involve the acoustic-articulatory process that produces interference due to acoustic similarity with recall.

The purpose of this experiment is to examine the effects of modality and similarity on Sternberg's (1967) context-recall paradigm. This task is very similar to the item-recognition task except that upon presentation of the test item, *S* must name the next item in the memory set that follows the test item. In the item-recognition task, *S* must say whether or not the test item is a member of the memory set. We expect a large effect of acoustic similarity and no effect of modality if, as Chase and Calfee (1969) suggest, a recall paradigm is the critical factor in Conrad's (1964) acoustic confusability effect.

Method.—Each trial of the context-recall paradigm consisted of a memorizing phase followed by a test phase. During the memorizing phase, *S* was presented with a list of three to six letters of the alphabet. When *S* indicated that he had memorized the list, the test phase followed with the presentation of a single test letter from the list. The *S*'s task was then to name, as quickly as possible and without errors, the letter following the test letter in the memorized list. If the test letter was the last letter of the list, then the response was the first letter of the list. The *S*'s latency was measured from the onset of the test letter to the vocal response.

Lists and test letters were presented either visually or aurally. During visual presentation, the memory set (typed in elite capitals) was seen on the top of a 5 × 8 card. Auditory presentations of the list consisted of *E* repeating the memory set, at approximately 2 letters/sec, with a 1-sec. pause at the end of the list.

The *S* took as much time as he wanted to memorize the lists. For auditory presentations, the number of repetitions of the memory set varied anywhere from 1 to as many as 10, depending upon *S*, size of the memory set, and the type of material. For visual presentations, *Ss* used less time, but again the time depended upon *S*, the size of the memory set, and the type of material.

If a visual test followed, *S* pressed a button, looked into a Polymetric two-channel tachistoscope, and 1 sec. later the test letter appeared and the timer started. At a viewing distance of 39.6 cm., the test letter, as well as letters in the memory set, subtended a visual angle of about 24′ vertically. If an auditory test followed, *S* said "ready" and looked into the tachistoscope. The *E* then said "ready," paused about 1 sec., and then said the test letter, which started the timer and presented a blank field in the tachistoscope.

The three types of material were neutral letters (A, D, H, I, M, Q, Y, Z), visually confusable letters (B, C, D, G, O, Q, R, U), and acoustically confusable letters (B, C, D, E, P, T, V, Z). Visual similarity was defined by an a priori overlap measure and acoustic confusability was based on Conrad's (1964) confusion data. The sets were closely equated for naming time and frequency of occurrence in the English language.

---

[1] This work was supported by Public Health Service Research Grant MH-07722 from the National Institute of Mental Health. Parts of this research were presented at the meeting of the Midwestern Psychological Association, Cleveland, May 1972.

[2] Requests for reprints should be sent to Michelene T. H. Chi, Department of Psychology, Carnegie-Mellon University, Schenley Park, Pittsburgh, Pennsylvania 15213.

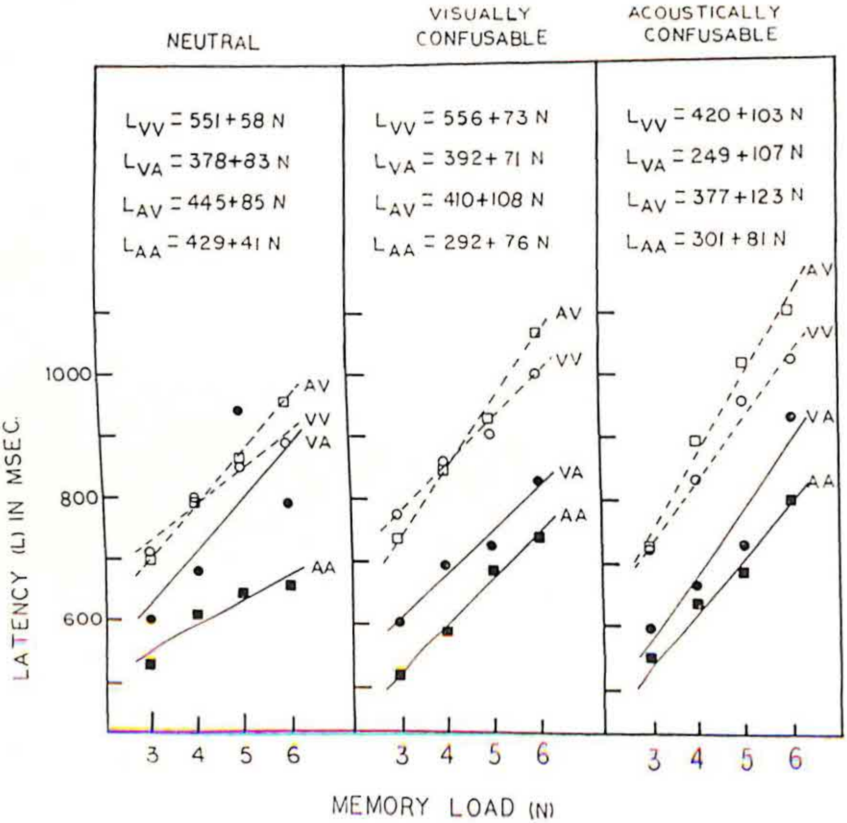


FIG. 1. Mean latency as a function of memory load for each similarity condition and each modality condition: Auditory list–auditory test (AA), auditory list–visual test (AV), visual list–auditory test (VA), and visual list–visual test (VV).

Eight *Ss* served in all four combinations of auditory and visual memory set and test letter presentations: Visual list–visual test (VV), visual list–auditory test (VA), auditory list–visual test (AV), and auditory list–auditory test (AA). The order of these four conditions was counterbalanced across *Ss* with a 4 × 4 Latin square.

For each of the four sessions, there were 192 trials plus 12 lead-in trials. The 192 trials consisted of 64 trials for each memory set size (3, 4, 5, 6) within each type of material. Each serial position of the test item was tested approximately equally often. The order of the 192 trials was randomized for each *S* and each session. Each session lasted about 1½ hr., there was a short break after 96 trials, and only one session was run per day.

Results.—Reaction time was a linear function of memory load under all modality and similarity conditions, the linearity accounted for 99.5% of the variance due to memory load, and there was no significant lack of fit, F (2, 21) < 1. Finally, there were strong serial position effects (all ps < .001) with the latency generally increasing linearly from Serial Position 1 onward, suggesting a serial self-terminating search process. All of these results are in accord with the original findings of Sternberg (1967). Hence, we use the slope of the line as a measure of the memory search rate. With a self-terminating search, the memory search rate is twice the slope of the regression line since *Ss* search half of the list on the average before finding the test item. The average slope of 84 msec. in this experiment is interpreted as a memory search rate of 168 msec/item. The intercept and slope parameters are shown in Fig. 1 for the modality and similarity conditions.

The most important result of the present experiment was that the search rates were faster by 48 msec/item when the memory set and test item were presented in the same mode (AA, VV) than in opposite modes (AV, VA); this contrast accounted for 96% of the variance due to the Memory Set Modality × Test Item Modality interaction on the search rates, F (1, 7) = 86, p < .001. This result is very similar to that reported by Chase and Calfee (1969) for recognition latencies.

The basic finding involving similarity was that search rates were slower through acoustically confusable letters than through visually confusable or neutral letters (58 msec/item average). This contrast accounted for 83% of the variance in search rates due to similarity, F (1, 14) = 20.2, p < .001. The 30-msec. difference between neutral and visually confusable search rates was not statistically significant, F (1, 14) = 4.1. Again, this

TABLE 1

ERROR PERCENTAGES FOR EACH LEVEL OF MEMORY LOAD, TYPE OF SIMILARITY, AND SERIAL POSITION

Memory load	Serial position	Similarity			
		Neutral	Visual	Acoustic	Average
3	1	2.6	2.1	2.1	2.3
	2	3.1	5.6	2.5	3.8
	3	3.1	10.6	3.8	5.8
	Average	2.9	6.1	2.8	4.0
4	1	7.0	4.7	2.3	4.7
	2	6.2	11.7	3.1	7.0
	3	3.9	1.6	6.2	3.9
	4	3.1	10.2	7.8	7.0
	Average	5.1	7.0	4.9	5.7
5	1	2.3	5.5	7.0	4.9
	2	12.5	5.2	29.2	15.6
	3	9.4	7.3	11.5	9.4
	4	11.5	9.4	8.3	9.7
	5	4.2	15.6	5.2	8.3
	Average	8.0	8.6	12.2	9.6
6	1	9.4	10.9	9.4	9.8
	2	15.6	17.2	29.2	21.1
	3	8.3	13.5	33.3	18.4
	4	7.3	6.2	24.0	12.5
	5	6.2	12.5	6.2	8.9
	6	4.7	8.3	17.2	9.8
	Average	8.6	11.4	19.9	13.4

result is similar to that reported by Chase and Calfee (1969) for recognition latencies.

There were no interactions involving modalities and similarities, which is again consistent with Chase and Calfee (1969).

In the present experiment, the intercept for auditory test items was 120 msec. lower than for visual test items, a result almost identical to Chase and Calfee's (1969) results for recognition latencies (118 msec.). This contrast accounted for 45% of the variance in the intercepts. Although the variance in the intercepts was too large to establish the reliability of this effect, the main effect of test-item modality was highly significant ($p < .001$).

There were also some artifacts in the intercepts that should be pointed out. For example, the intercept for acoustically confusable materials was 75 msec. lower than for the other two conditions. And for the modality conditions, equating for the modality of the test item, it can be seen in Fig. 1 that steeper slopes were generally associated with lower intercepts (e.g., AV vs. VV). This is due to the fact that the cross-modality effect and acoustic similarity had their effects at larger memory loads, and latencies are about the same for all conditions for a memory load of three items (with the exception of the test-item modality, which is a real effect), and extrapolating back to zero distorts the intercepts. Although we are mainly concerned with the slopes in this study, two points should be made. First, no psychological significance should be attached to the intercept effects in this case. And second, it is important to exercise caution in interpreting the intercept in the serial search model; the model in fact needs to be extended to get a clear estimate of the intercept.

Table 1 illustrates several important results concerning errors: (*a*) Errors increased with memory load ($p < .001$). (*b*) For memory loads of five and six items, there was a serial position effect ($p < .001$, in both cases). This serial position effect resembled the classical serial position curve, except that the last item in the list showed more errors than normal, probably because the item was associated with the first item of the list. (*c*) There were more errors for acoustically similar material, but only for memory loads of five and six items ($p < .001$), and then only for items in the middle of the list ($p < .001$).

It is interesting to note that with a list of six acoustically confusable letters, *S*s averaged as much as 33% errors in the middle of the list even when they were given as much time as they needed to memorize the lists (*S*s generally did take longer on the larger acoustically confusable lists). The error data, then, are in agreement with Conrad (1964) that acoustic similarity is a powerful cause of forgetting in short-term memory.

Discussion.—The most interesting result of the present experiment is the faster search rates when the memory set and test item are presented in the same modality, a result quite similar to that found by Chase and Calfee (1969) for recognition memory. This finding agrees with the hypothesis that specific sensory information is also stored in short-term memory, making possible a faster physical match when the memory set and test item are presented in the same mode. When different modes are used, the comparison process presumably must involve a higher level of representation in which the auditory and visual exemplars are identical. Thus, in both the recognition and recall paradigms, the search process can apparently be based on a direct comparison of sensory features.

The findings on similarity, however, pose some problems for the sensory coding hypothesis since there were no Modality × Similarity interactions in either the present experiment using a recall paradigm or in the earlier experiments (Chase & Calfee, 1969) using a recognition paradigm. According to the sensory coding hypothesis, visual confusability should influence the visual code and acoustic confusability should influence the auditory code. The only significant effect of similarity, however, was a slower search rate through acoustically confusable material regardless of list or test modality. These results provide some support for Conrad's (1964) acoustic coding hypothesis. An alternative hypothesis, however, is that acoustic confusability has its effect on the retention process prior to the search, as evidenced by the greater error rates. That is, acoustic similarity may weaken the memory of the list, and search through these weaker memory traces is slower.

Another result of interest is the 120-msec. advantage of auditory over visual test items. At least part of this difference can be attributed to sensory processes in the sense that auditory latencies are faster than visual latencies by about 50 msec. for simple stimuli such as lights and tones. Although we don't know why this modality effect is larger for more complex stimuli, we can surmise that modality of the test item has its effect on a stage of processing that is independent of the comparison process, perhaps at an earlier stage, since this modality effect does not interact with variables affecting the comparison process.

Another question of concern is how well Sternberg's (1967) serial self-terminating model accounts for the data in the present experiment vis-à-vis some kind of parallel search model. At first glance, the increasing latencies with serial position seem to be strong evidence for such a serial self-terminating model. However, there is a parallel self-terminating model which also makes these same predictions. Suppose it is assumed that only a primacy effect occurs in these data so that the associative strengths decrease across serial positions. If it is further assumed that (*a*) the time to compare the test item against a member of the memory set is proportional to the serial position, (*b*) all comparisons are carried out independently in parallel, and (*c*) the search is terminated as soon as the test item is located in the list, then this model makes exactly the same predictions as Sternberg's model and, in addition, the serial position effect is explicitly predicted. However, we must reject this model in favor of Sternberg's because the error data of Table 1 disconfirm the assumption that associative strengths decrease over serial positions. Rather, associative strengths resemble the classic serial position curve. We conclude, therefore, that context recall involves a serial self-terminating search process.

REFERENCES

Chase, W. G., & Calfee, R. C. Modality and similarity effects in short-term recognition memory. *Journal of Experimental Psychology*, 1969, **81**, 510–514.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

Posner, M. I., & Keele, S. W. Decay of visual information from a single letter. *Science*, 1967, **158**, 137–139.

Sternberg, S. High speed scanning in human memory. *Science*, 1966, **153**, 652–654.

Sternberg, S. Retrieval of contextual information from memory. *Psychonomic Science*, 1967, **8**, 55–56.

(Received February 28, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 223–225

SEMANTIC AND SYNTACTIC CONSTRAINTS ON FREE-RECALL LEARNING OF SENTENTIAL MATERIAL[1]

VERNE R. BACHARACH[2]

Bureau of Child Research, University of Kansas

The present experiment was designed to investigate effects of semantic content and phrase structure on multitrial free recall. Levels of semantic content were produced by diagonalizing easily understood sentences and scrambling words within each of the diagonalized sentences. Phrase structure was manipulated by presenting the material in either the active or passive voice. Four levels of compressed speech were orthogonally varied with the other factors. It was hypothesized that semantic content and phrase structure would interact. Specific predictions about the nature of the interaction were made following an examination of previous research. It was also assumed that the magnitude of the semantic content and the phrase structure effects would increase as degree of speech compression increased. The whole sentence correct response data and the transformational error data supported the hypothesis that certain semantic and syntactic features of sentences interact. The results partially supported the hypothesis regarding compressed speech.

The effects of syntactic constraints and semantic variables on various behavioral indices have been reported by numerous researchers (Bacharach & Kellas, 1971; Howe, 1970; Marks & Miller, 1964). An implicit assumption made in most of these experiments is that semantic and syntactic features of sentences do not interact. This assumption has been used to justify manipulations which might otherwise be considered confounded. For example, manipulation of transformational complexity is always associated with changes in either or both semantic features or phrase structure characteristics of sentences. Recent experimental evidence suggests that the assumption of semantic and syntactic independence is questionable (Howe, 1970; Rosenberg, 1968; Wearing, 1970).

The present experiment was designed to test the hypothesis that the effects of certain semantic and syntactic features of sentences will interact in a multitrial free recall task. Semantics was manipulated by varying the degree of semantic content of the sentences. Semantic content was defined with a diagonalization procedure used by Marks and Miller (1964) and others. The procedure produces "anomalous" sentences that are composed of words from the original, easily understood sentences. In addition to the original and diagonalized sentences, some *S*s in the present experiment were asked to learn strings of words that were grammatically unstructured. These strings were produced by randomly scrambling the words within each of the diagonalized sentences. Because these strings were grammatically unstructured, it was hypothesized that there would be no phrase structure differences between them. Syntax was manipulated by varying the phrase structure characteristics of the sentences. Active and passive forms of the same sentences were employed. Manipulating syntax in this way has the advantage of producing phrase structure differences without changing the content words used in the sentences.

The semantic content and the phrase structure factors were orthogonally varied over trials with four levels of compressed speech. Compressed speech was used in this experiment to systematically degenerate the speech signal. It was reasoned that compressed speech might amplify small magnitude effects which would otherwise go undetected. Additionally, it should be possible to determine if semantic and syntactic constraints are differentially affected by levels of speech compression.

Method.—The *S*s were 168 undergraduate students from the University of Kansas, who were fulfilling a general psychology course requirement.

The *S*s were seated at a table facing a partition. A button and a telegraph key were located on the table in easy reach of the *S*s. The *S*s wore stereo headphones and listened to the tape-recorded material played on a stereo tape recorder. A buzzer controlled by an ATC timer was used to signal *S*s to begin responding. The *S*s' responses were recorded on a tape recorder.

The stimulus materials consisted of tape-recorded strings of words (hereafter referred to as sentences). There were three types of sentences: High Semantic Content (HSC) sentences such as "The dying horses kicked the trainers," Low Semantic Content (LSC) sentences such as "The torn snakes cluttered the trainers," and No Semantic Content (NSC) sentences such as "Torn cluttered the snakes trainers the." All sentences were six words long with two function words and four content words in each sentence.

Half of each of the different semantic content sentences were in the active voice and half in the

---

[1] The present research was supported by United States Public Health Service Grant HD 00870 and HD 00183 to the Bureau of Child Research, University of Kansas. This research is based on a doctoral dissertation submitted to the University of Kansas in partial fulfillment of the requirements for a PhD degree. The author wishes to acknowledge George Kellas and Edward Wike for their helpful suggestions and comments.

[2] Requests for reprints should be sent to Verne R. Bacharach, Bureau of Child Research, University of Kansas, Lawrence, Kansas 66044.

TABLE 1

Mean Whole Sentence Correct Recall Scores for Active and Passive Sentences as a Function of Semantic Content

Phrase structure	Semantic content		
	HSC	LSC	NSC
Active	2.97	2.13	.83
Passive	3.10	2.04	.77

Note.—Each point represents an average score collapsed over trials and lists.

passive voice. An *S* was asked to memorize eight sentences, four active and four passive, of one semantic content type. To avoid semantic overlap, two lists of sentences were produced with the active sentences in one list transformed to passive sentences in the other list and vice versa.

The groups of eight sentences were tape recorded in a random order, with the restriction that no more than two sentences with the same syntactic form would occur in sequence. Twenty random orders of the eight sentences for the three sentence types within each list were recorded. The materials were recorded at four different levels of speech compression. The four levels of compressions were 0%, 15%, 30%, and 45%. The deletion sample size was approximately 35 msec. under each compression condition. The sentences were taped and played at $3\frac{3}{4}$ ips.

There was a 2-sec. intersentence interval within a group of eight sentences, and a 10-sec. intergroup interval between each group of eight sentences.

The *S*s were told that they would hear eight strings of words and that they would be given a number of trials to learn the material. The *S*s were told that the strings of words would be presented in a random order on each trial, and that order of recall was not important. However, the instructions stressed the importance of ordered recall of the words within each string.

After each presentation of the eight sentences, a buzzer sounded signaling *S*s to begin recall. After recalling a sentence or any part of a sentence, *S*s pressed the button on the table to indicate response termination. After recalling as many of the strings as they could remember, *S*s pressed the telegraph key. By pressing the telegraph key, *S*s simultaneously turned off the recorder recording their responses and turned on the recorder used to present the taped material. The recall interval was unlimited, although no *S* exceeded 2 min.

The learning criterion was 2 consecutive errorless trials or 20 trials, whichever came first. Pilot data showed that approximately 96% of the pilot *S*s given the HSC and the LSC lists reached correct recall learning criterion. Pilot *S*s asked to learn the NSC material were correctly recalling approximately 23% of the sentences after 20 trials.

Results.—Whole sentence correct response scores were analyzed. A response was scored as correct if all content words in the sentence were correctly recalled in their presented order and sentence voice was not changed. A change in number (i.e., singular-plural) was not considered as an error. Substitution of function words was permissible to the extent that the substituted word was from the same form class as the original word in the sentence.

A 4 × 3 × 2 × 2 × 20 analysis of variance was performed on these data. Compressed speech (4 levels), semantic content (3 levels), and list (2 levels) were between-groups factors. Phrase structure (2 levels) and trials (20 levels) were within-*S* factors. Following the suggestion of Wike (1971), conservative *F* tests were conducted on the obtained *F* values involving within-*S* factors. The results of this analysis showed reliable main effects for compressed speech, $F\ (3,\ 144) = 2.72,\ p < .05$; semantic content, $F\ (2,\ 144) = 245.17,\ p < .001$; and trials, $F\ (1,\ 177) = 341.37,\ p < .001$. The Compressed Speech × Semantic Content interaction was also significant, $F\ (6,\ 144) = 2.25,\ p < .05$. There were no effects of compressed speech under the HSC and NSC conditions. With the LSC material, correct responding increased as compression ratios decreased.

Most importantly, the overall analysis showed that the Semantic Content × List × Phrase Structure × Trials interaction was statistically reliable, $F\ (2,\ 144) = 4.22,\ p < .05$. This interaction was examined further by analyzing the effects of the Semantic Content × Phrase Structure × Trials interaction as a function of each list condition. A reliable three-factor interaction was obtained in both analyses: List 1, $F\ (38,\ 1529) = 2.33$; List 2, $F\ (38,\ 1539) = 2.57$; both $ps < .01$. The list and trial effects had no theoretical import. List effects were due to magnitude differences on different blocks of trials. The trial effects were accounted for by greater Semantic Content × Phrase Structure differences as the number of trials increased. Table 1 shows the mean whole sentence correct recall scores for the phrase structure conditions as a function of semantic content.

Tukey *b* multiple comparisons were used to analyze the differences between active and passive sentences for each level of semantic content in those blocks of trials in which an interaction was found. The results of these comparisons showed that (*a*) there were no reliable differences between active and passive sentence recall when HSC material was used; (*b*) active sentence recall was significantly better ($\alpha = .05$) than passive sentence recall with the LSC material; and (*c*) recall of NSC active sentences was superior to recall of NSC passive sentences. There was one exception to these results. In one block of trials, passive sentence recall was superior to active sentence recall in the HSC condition. This difference was eliminated as a function of LSC and NSC. Nevertheless, the direction of the interaction in the exceptional case was the same as those found in the other blocks of trials.

A response was scored as a transformational error if *S* recalled the sentence correctly but changed sentence voice. There were a total of 214 transformational errors. Of these errors, 97% occurred in the HSC condition. Transformational errors did not appear to be systematically related to any of the other factors in this experiment.

An intrusion error occurred when *S* substituted an incorrect word for one of the content words in a sentence that was otherwise recalled correctly. The frequency of intrusion errors for each type of content word was tabulated.

The intrusion error data showed that the most frequent errors were verb intrusions. Of the 1,546 intrusion errors, 56% were verb intrusions, 23% adjective intrusions, 12% noun$_2$ (second noun in a sentence), and 8% noun$_1$ (first noun in a sentence) intrusion errors. There appeared to be a large effect for semantic content. Of the errors, 46% occurred in the LSC condition, while 26% occurred in the HSC condition. The NSC condition accounted for 28% of the intrusion errors.

Discussion.—The present experiment was designed specifically to test the hypothesis that semantic content and phrase structure have interactive effects on the learning of sentential material. The four-factor interaction involving semantic content and phrase structure clearly showed that the effects of phrase structure depend on the degree of semantic content of the material to be learned. As a general rule, it appears that the magnitude of phrase structure effects increases as semantic content decreases. In the present experiment, active–passive differences increased as semantic complexity decreased. Although in the present experiment this effect could only be demonstrated for active and passive constructions, the results of earlier research also tend to support such an hypothesis (Rosenberg, 1968; Wearing, 1970).

The transformational error data also support an interaction hypothesis. Transformational errors theoretically reflect mistakes in processing syntactic features of sentences. The transformational error data in this experiment clearly indicated that these types of errors were a function of the semantic features of the sentences. If transformational errors do represent errors in syntactic processing, then the probability of making a syntactic error increases as the semantic content of a sentence increases. It is as if *S*s paid less attention to the syntactic features of sentences that were easy to understand than they did to the syntactic features of difficult-to-understand sentences. Traditionally, it has been assumed that transformational errors increase as the transformational complexity of a sentence increases. The transformational error data in the present experiment indicate that for active and passive constructions, semantic features of the sentences are more important than transformational histories in determining the frequency of transformational errors.

One aspect of the proposed hypothesis was not supported by the whole sentence correct recall data. It had been assumed that recall of the NSC sentences would not be influenced by phrase structure. Although the NSC sentences were grammatically unorganized, there were phrase structure differences within the NSC condition. Because type of function word was the only difference between the "active" and "passive" NSC sentences, this finding was not anticipated. It appears that function words may not be a homogeneous grammatical set of items.

Finally, the nature of the interaction of compressed speech with semantic content indicates that semantics plays a greater role in the processing of speech compressed sentences than do the grammatical features of the material.

REFERENCES

Bacharach, V. R., & Kellas, G. Phrase versus base structure effects on short-term retention. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 171–175.

Howe, E. S. Transformation, associative uncertainty, and free recall of sentences. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 425–431.

Marks, L. E., & Miller, G. A. The role of semantic and syntactic constraints in the memorization of English sentences. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 1–5.

Rosenberg, S. Association and phrase structure in sentence recall. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 1077–1081.

Wearing, A. J. The storage of complex sentences. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 21–29.

Wike, E. L. *Data analysis*. Chicago: Aldine-Atherton, 1971

(Received March 2, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 226–228

FUNCTIONAL UNITS IN FREE RECALL[1]

JAMES FRITZEN AND NEAL F. JOHNSON[2]
Ohio State University

In 1968, E. Tulving and R. D. Patterson compared *S*s' free recall of lists which contained related words in the middle (M list) or at the end (E list) with the free recall of lists containing only unrelated words. For the E list, an unextended recency effect was taken as support for lack of unitization of related words in short-term memory. In the present experiment, an attempt was made to ensure unitization of the related words by having *S*s memorize them before hearing the list and then informing *S*s as to where those words would appear within the list. Slight evidence for an extended recency effect for the E list was nullified by an equivalent elevation in recall of the words preceding the related words in the M list. The results raised questions regarding Tulving and Patterson's index of unitization.

Tulving and Patterson (1968) had *S*s free recall lists of words which were either (*a*) all unrelated to one another (control or C list), (*b*) all unrelated except for four related words (e.g., north, east, south, and west) occurring at the end (E list), or (*c*) all unrelated except for four related words in the middle of the list (M list). There were more words recalled from the M list than the C list. However, when recall was scored by counting each unrelated word as a functional recall unit and the category as a unit, regardless of the number of words recalled from the category, the two conditions were equivalent. When the E list was scored in the same manner for functional units in recall, the mean was lower than for the other two lists.

Proceeding on the assumption that items from the end of the list are recalled from short-term or primary memory and other items are recovered from long-term or secondary memory, Tulving and Patterson (1968) argued that their data indicated that reorganization of information into higher order memory units occurred in secondary memory, but not in primary memory. (They equated the two memory systems with different retrieval mechanisms rather than differences in the nature of the storage.)

To further support their position, Tulving and Patterson (1968) pointed out that the recency effect was not extended across more end items when the last four were related than when they were not related. Their argument was that if the last four items were related, and they represented a single functional unit in primary memory, then those items should occupy no more space in primary memory than one unrelated word (e.g., the above set might be stored as *directions*). Therefore, recall of the fifth item from the end for the E list should be about the same as for the second from the last word in the C list, etc. Those results were not obtained, and recall for the fifth item from the end was about the same for the two types of list.

If *S*s were to memorize the four related words before the list was presented to them, and they were told exactly where those items would appear within the list, then according to the position presented by Tulving and Patterson (1968) these items should be unitized in secondary memory. Furthermore, their position would predict that if these preunitized items appeared at the end of the list, there should be an extended recency effect and the number of functional units recalled should be the same as for the C and M lists.

Other interpretations of the Tulving and Patterson (1968) results appear possible. It would seem that it would be necessary for *S*s to hear the four related words at the end of the list and store them in primary memory, before *S*s could detect that the items were a category. If they then recoded the items into a functional unit within primary memory, the initial storage of the items would still be necessary and there would be no extension of the recency effect, even though unitization occurred.

Another possibility is that the level of recall of the items at the end of the list may be a function of the time between the input and output of these items. Therefore, *S*s' ability to recall the fifth item from the last would be a function of the number of words which followed it during presentation. In that the fifth item from the last was followed by four other words in both the E and C lists, these two conditions should be equivalent in the level of recall of that item.

While the position adopted by Tulving and Patterson (1968) would predict an extended recency effect if the last four items in the list were a preunitized category, neither of the alternative interpretations presented here would make such a prediction. Both interpretations would predict that recall of an item near the end of a list would be a function of the number of items which followed it during presentation and not influenced by whether these following items also constituted a functional unit. The issue, then, is the extent to which the Tulving and Patter son method can be used to assess unitization in memory as they define it.

---

[1] The study reported here was supported by Grant MH08526 from the National Institute of Mental Health, United States Public Health Service, and sponsored, in part, by the National Science Foundation through Grant GN534.1 from the Office of Science Information Service to the Computer and Information Science Research Center, Ohio State University.

[2] Requests for reprints should be sent to Neal F. Johnson, Department of Psychology, Ohio State University, Columbus, Ohio 43210.

Method.—In addition to Tulving and Patterson's (1968) C, M, and E lists, two types of instructions were used prior to the M and the E lists. The not-informed (NI) instructions required *Ss* to recall as many words as they could in any order. The informed (I) instructions provided *Ss* with the four related words to be used, and indicated their locus in the list. Instructions for the C lists were always of the NI type. Thus, there were five conditions: C-NI, M-NI, M-I, E-NI, and E-I. Each *S* served in each condition twice.

Fifty-five *Ss* served in the experiment as part of an introductory psychology course requirement. The *Ss* were run in eight groups of either 6 or 7 *Ss* each. At the beginning of the experiment, *Ss* were fully instructed regarding the various types of lists they would be presented. The order of the 10 lists was determined by a table of random numbers for each of the groups.

All lists were 16 words in length. Unrelated words were drawn from a pool of nouns or words that could be used as nouns, having a Thorndike-Lorge count of A or AA. The category pool consisted of eight exhaustive and four nonexhaustive categories selected from Cohen (1963). The categories were assigned randomly to conditions within each of the eight groups of *Ss*. Some of the exhaustive category members had a frequency count of less than A. Related words occupied Positions 6–9 and 13–16 in the M and E lists, respectively.

Words were presented orally at the rate of 2 sec. per word. Recall occurred immediately after presentation, and *Ss* wrote the words on blank sheets of paper. Sixty seconds were allowed for recall, after which the sheets were placed face down.

Before each M-I and E-I list, *Ss* copied down the four related words to be used in that list and were given 30 sec. to memorize them. They were not allowed to see the words at the time of recall. The *Ss* were told the words would appear as the last four words in the list (E-I) or in a group in the middle of the list (M-I), depending upon the condition that list represented.

Results.—The *Ss'* recall protocols were scored for functional units by recording one point for each unrelated word recalled and one point if at least one related word was recalled. The mean number of functional units recalled for each condition were: C = 6.79, M-NI = 6.17, M-I = 7.24, E-NI = 5.19, and E-I = 5.98. (The mean number of words recalled for the five conditions were: 6.79, 8.73, 10.27, 8.02, and 8.95.)

The functional unit data for the four experimental conditions were subjected to a 2 × 2 analysis of variance. The effect of locus of the related items was significant beyond the .001 level, F (1, 54) = 51.58, as was the effect of instructions, F (1, 54) = 46.32. The Locus × Instruction interaction was not significant, $F < 1.00$. These data support Tulving and Patterson (1968) to the extent that the M lists yielded higher functional unit recall than E lists. In addition, prior learning of the related items also elevated *Ss'* recall level, but the lack of an interaction would suggest that the amount of improvement was independent of the location of the related items.

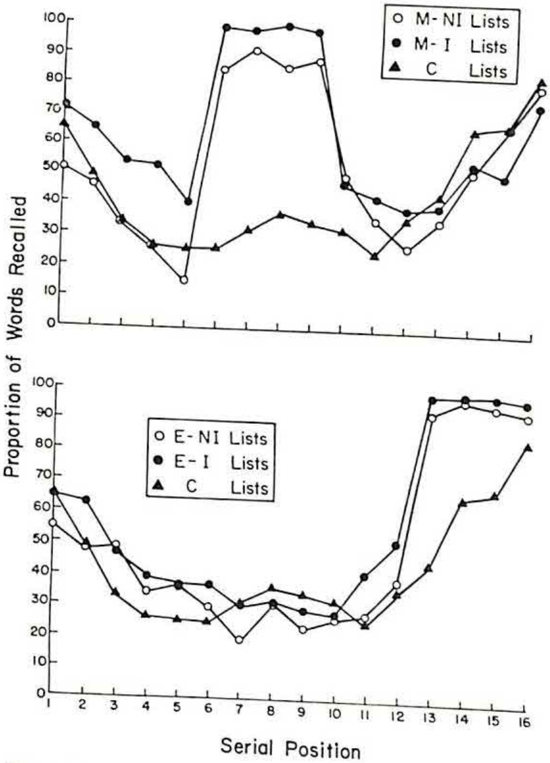


FIG. 1. Proportion of words recalled as a function of serial position. (The C list in both graphs is the same.)

The comparison between the C and the E-NI lists was significant, t (54) = 7.55, $p < .001$, replicating the Tulving and Patterson (1968) result. However, the difference between C and M-NI also was reliable, t (54) = 2.47, $p < .01$, which is not in accord with the results reported by Tulving and Patterson. The recall of the E-I lists also was below that of the C lists, t (54) = 3.19, $p < .01$.

The serial position curves for the five conditions in the experiment are given in Fig. 1. To examine the extent to which the recency effect was extended, *Ss* in the experimental conditions were scored for the recall of the two items immediately preceding the related items. Each *S*'s score varied from 0 to 4. As with Tulving and Patterson (1968), the recall of these items for the E-NI list did not exceed that for the C lists, t (54) = .82, with the means being 1.35 and 1.22. Mean recall of these items for the E-I list was 1.84, which was significantly above that for the E-NI list, t (54) = 2.23, $p < .05$, and the C list, t (54) = 3.26, $p < .005$.

The same pattern of results was found for the two unrelated items in the M lists, but the difference was slightly greater. For C, M-NI, and M-I, the means were 1.02, .84, and 1.84, respectively. The difference between C and M-NI was not significant,

$t = 1.00$, while the difference between M-NI and M-I was significant, $t\ (54) = 6.66$, $p < .001$, as was the difference between the C and the M-I lists, $t\ (54) = 4.77$, $p < .001$.

Recall of these items for the four experimental conditions was analyzed in a 2 × 2 analysis of variance. The effect of instructions was significant beyond the .001 level, $F\ (1, 54) = 40.21$, but neither the effect of locus, $F\ (1, 54) = 3.15$, nor the interaction, $F\ (1, 54) = 3.09$, was significant.

Discussion.—The major issue examined by the present experiment was not Tulving and Patterson's (1968) conclusion that organization did not occur in primary memory, but rather the extent to which their analysis of the functional unit recall and the recency effect can answer that particular question.

Regarding functional unit recall, the present data did not support the earlier finding that the number of functional units recalled was independent of their size (note—there were methodological differences between the studies). Furthermore, when *S*s learned the related items before the lists were presented, there was an overall increase in the number of functional units recalled, but there was no differential effect in favor of the E list. If the original difference between the E and M lists was attributable to the fact that the related items were not unitized in the E list, that differential improvement should have occurred.

The difference in functional unit recall between the E and M lists obtained by Tulving and Patterson (1968), as well as that reported here, could be the result of an artifact. The recall level for the words at the end of the list is generally higher than for words in the middle of the list. If, for the C lists, *S*s recalled an average of only one word from four positions in the middle of the list, then *S*s' recall score would be about the same regardless of whether they were given one point for every word recalled from those positions or whether they were given one point for recalling at least one of the words from those positions, regardless of how many other words were produced. On the other hand, if *S*s recalled an average of two words from the last four positions, then the two scoring procedures would not yield equivalent scores. The *S*s would have recalled two words, but only one functional scoring unit. Therefore, it is possible that the difference between the E and M lists may have been a function of the scoring procedures, and the shape of the serial position curve, rather than any critical difference between the two list types.

The elevated recall of the two items preceding the category in the E-I list might be taken to indicate an extended recency effect when the end items were preunitized. However, the same effect occurred for the two items preceding the category in the M-I list where an extended recency effect is not a reasonable interpretation. In both cases, rehearsal of the just-presented items while the category was being read seems to be more in accord with the data.

In summary, Tulving and Patterson (1968) attribute the recall differences between their E and M lists to the fact that the related items were unitized in the M lists but not in the E lists. The present experiment suggests the same recall differences are obtained when the related items are unitized in both types of lists.

REFERENCES

COHEN, B. H. An investigation of recoding in free recall. *Journal of Experimental Psychology*, 1963, **65**, 368–376.

TULVING, E., & PATTERSON, R. D. Functional units and retrieval processes in free recall. *Journal of Experimental Psychology*, 1968, **77**, 239–248.

(Received March 3, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 229–231

EFFECTS OF PARTIAL REINFORCEMENT IN ONE OR BOTH GOAL BOXES OF A DOUBLE ALLEYWAY[1]

JOSEPH A. SGRO,[2] WILLIAM B. PAVLIK, JOHN R. SHOWALTER,[3] AND NEIL H. COHN

Virginia Polytechnic Institute and State University

Forty-eight albino rats received 64 trials in a double alleyway with either 50% or 100% reinforcement in the first goal box and 50% or 100% reinforcement in the second goal box. All *S*s were then shifted to 100% reinforcement in Goal Box 1 and 0% in Goal Box 2 for 24 additional trials. Preshift results indicated that second alleyway speeds were inversely related to both Goal Box 1 and Goal Box 2 reward percentage. The frustration effect occurred with the effect being greater for running speeds of *S*s which received 100% reinforcement in Goal Box 2. Postshift Alley 2 data indicated that resistance to extinction was inversely related to Goal Box 2 reward percentage. Goal Box 1 preshift training did not affect resistance to extinction in Alley 2.

Studies which have manipulated partial reinforcement in the first goal box of a double alleyway have shown that second alleyway speeds are faster following nonreward than following reward. This finding has been designated the "frustration effect" (FE) and has been assumed to reflect the motivational properties of frustrative nonreward (Amsel, 1958). A review of the literature has revealed that in demonstrating FE, second-goal-box reward conditions, in general, have remained constant. Although studies have investigated the FE as a function of reward magnitude (Krippner, Endsley, & Tacker, 1967) and reward homogeneity (Levy & Seward, 1969) in Goal Box 2, no evidence is available concerning the influence on the FE of partial reinforcement in Goal Box 2.

The present study was designed to accomplish two objectives. First, it was of interest to investigate the effect of partial reinforcement in *both* Goal Box 1 and Goal Box 2 on Alley 2 performance. This manipulation offered an opportunity to examine the magnitude of the FE as a function of partial reinforcement in Goal Box 2. To this end, four groups received a factorial combination of partial reinforcement in Goal Box 1 (50% vs. 100%) and Goal Box 2 (50% vs. 100%) for 64 trials. The second purpose of the present study was to determine the extent to which partial reinforcement in either or both goal boxes would affect resistance to extinction in Alley 2. Since it has been shown that partial reinforcement increases resistance to extinction in a single alleyway (Goodrich, 1959), it was expected that partial reinforcement in Goal Box 2 would produce greater resistance to extinction in Alley 2. Within frustration theory, greater resistance to extinction is attributed to the prior conditioning of approach responses to the cues of frustrative nonreward. Of major concern, however, was whether partial reinforcement in Goal Box 1 would affect resistance to extinction in Alley 2. The theoretical consideration was whether the frustration responses elicited in Goal Box 1 would perseverate and become conditioned to stimuli in Alley 2 thereby affecting subsequent resistance to extinction. The four groups were thus shifted to 100% reinforcement in Goal Box 1 and 0% in Goal Box 2 for 24 trials.

Method.—The *S*s were 48 experimentally naive male albino rats obtained from Dublin Laboratories, Dublin, Virginia, and were 90–100 days old at the beginning of the experiment.

The apparatus was an L-shaped double alleyway consisting of a first start box, a first alley, a first goal box, a second alley, and a second goal box. The first goal box also served as the start box for Alley 2. Guillotine doors separated the start box from Alley 1, Alley 1 from Goal Box 1, Goal Box 1 from Alley 2, and Alley 2 from Goal Box 2. The dimensions and construction of the apparatus were identical to that employed by Sgro, Glotfelty, and & Podlesni (1969).

Start times for Alley 2 were measured from the opening of the Goal Box 1 door leading into Alley 2 until the breaking of a beam 7.62 cm. distant. Running times for Alley 2 were measured from the breaking of the latter photobeam to the interruption of a second beam 182.88 cm. distant.

The *S*s were randomly assigned to four equal preshift groups ($n = 12$) which comprised a 2 × 2 factorial design. There were 2 levels of percentage of Goal Box 1 reward (50% vs. 100%) and 2 levels of percentage of Goal Box 2 reward (50% vs. 100%). After 64 preshift trials, all *S*s were shifted to continuous reward in Goal Box 1 and nonreward in Goal Box 2 for 24 trials. The four preshift groups were designated 100–100, 100–50, 50–100, and 50–50 in which the first number indicates the percentage of reward received in Goal Box 1 and the second number refers to the percentage of reward in Goal Box 2. During the preshift session, the 50% Goal Box 1 groups received the following daily cycles of reward (R) and nonreward (F) in Goal Box 1: RRFF, FFRR, RFFR, FRRF, FRFR, RFRF. The 50% Goal Box 2 groups received the following daily cycles in Goal Box 2: RFRF, FRFR, FRRF, RFFR, FFRR, and RRFF. These patterns were repeated over the entire experi-

---

[1] An expanded version of this paper was presented to the Southwestern Psychological Association in St. Louis, April 1970.

[2] Requests for reprints should be sent to Joseph A. Sgro, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

[3] Now at Purdue University, Lafayette, Indiana.

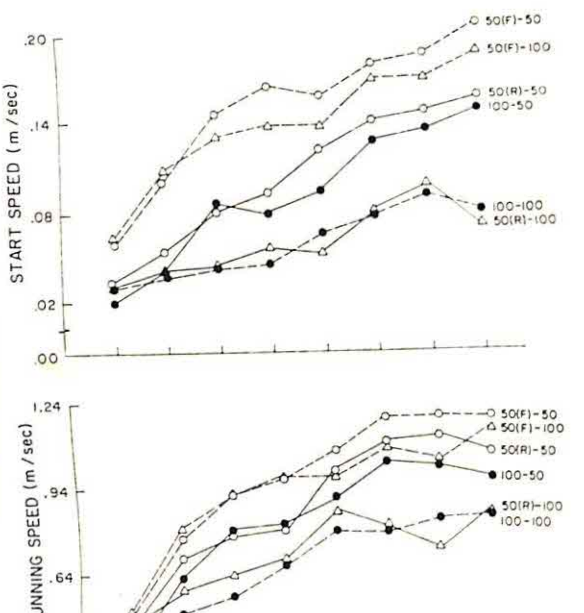


FIG. 1. Mean Alley 2 start and running speeds as a function of trial blocks during the preshift phase. (Trial blocks for Groups 100–50 and 100–100 are plotted in blocks of eight, whereas trial blocks for the within-*S* conditions of Groups 50–50 and 50–100 are plotted in blocks of four.)

ment. On rewarded trials in either goal box, *S* was provided with a 20-sec. access to six Noyes 45-mg. food pellets and on nonrewarded trials, *S* was given a 20-sec. confinement in the appropriate goal box. The food pellets were available to *S* in a plastic cup located in front of the guillotine door leading to Alley 2 and also in a small bottle cap located at the end wall of Goal Box 2.

Fourteen days prior to the first preshift trial, *Ss* were placed on a $22\frac{1}{2}$-hr. food-deprivation schedule. During pretraining and throughout the course of the experiment, *S* was allowed $1\frac{1}{2}$-hr. access to food in the home cage after the daily session. During Days 1–8, each *S* was gentled for approximately 3 min. daily. On Days 9–11, each *S* was allowed to explore the entire apparatus for 2 min. After exploration, *S* was given access to scattered 45-mg. Noyes pellets in a neutral holding cage before being returned to the home cage. On Days 12 and 13, all *Ss* were given one pretraining trial with six food pellets available in Goal Box 1 and Goal Box 2, and, on Day 14, *Ss* were given two pretraining trials with six food pellets in Goal Box 1 and Goal Box 2. On Preshift Day 1 and throughout the entire experiment, *S* received four daily trials with an intertrial interval of approximately 20–30 sec. During the intertrial interval, *S* was placed in a neutral holding cage.

A typical trial consisted of placing *S* into the start box. After 5 sec., *E* raised the guillotine door permitting *S* to traverse Alley 1. Upon entering Goal Box 1, the guillotine door was closed to prevent retracing and *S* was allowed to eat the available pellets. After 20 sec., the door leading to Alley 2 was opened and *S* was allowed to traverse Alley 2.

Results and discussion.—Starting and running time measures for each trial were converted into speed scores (m/sec), and all analyses were performed using the converted measures.

Figure 1 presents preshift Alley 2 start and running speeds as a function of trials in blocks of eight. In Fig. 1, separate means for four R trials and four F trials within each block are plotted for Groups 50–50 and 50–100. Analyses of variance performed on Alley 2 start and running speeds for Groups 50–50 and 50–100 over the eight preshift trial blocks indicated that: (*a*) The FE effect was significant for both the start and running measures, F (1, 22) = 50.84 and 48.88, respectively, $p < .01$; (*b*) the FE × Blocks interaction was statistically significant for both the start and running measures, F (7, 154) = 2.73 and 3.37, respectively, $p < .05$; (*c*) the Groups × FE interaction was statistically reliable for running speeds, F (1, 22) = 7.13, $p < .05$, but not for start speeds, F (1, 22) = 2.78, $p > .05$; (*d*) the Blocks × Groups interaction was statistically significant for start speeds, F (7, 154) = 3.15, $p < .01$, and running speeds, F (7, 154) = 2.30, $p < .05$; and (*e*) the blocks variable was significant beyond the .01 level for both the start and running measures, F (7, 154) = 42.91 and 51.09, respectively, $p < .01$. The Blocks × Groups × FE interaction failed to achieve statistical significance for start speeds, F (7, 154) = 1.03, $p > .05$, or running speeds, F (7, 154) = 1.12, $p > .05$.

Analyses of variance performed on start and running speeds over Trial Blocks 6–8 for Groups 50–50 and 50–100 revealed that: (*a*) The FE variable was significant for both start and running measures, F (1, 22) = 39.64 and 44.02, respectively, $p < .01$; (*b*) the FE × Groups interaction was significant for start speeds, F (1, 22) = 4.65, $p < .05$, and running speeds, F (1, 22) = 10.47, $p < .01$; and (*c*) Group 50–50 ran significantly faster than Group 50–100 over the start measure, F (1, 22) = 8.43, $p < .01$, and the running measure, F (1, 22) = 6.50, $p < .05$. Neither the blocks variable nor the interactions containing the blocks variable approached statistical significance (both *ps* > .05) for the start or running measures. Further analyses indicated that while all groups showed significant FE effects for both start and running speeds (all *ps* < .01), the significant Groups × FE interactions were due to the faster Alley 2 speeds exhibited by Group 50–50 following R trials in Goal Box 1 for the start measure, F (1, 22) = 10.19, $p<.01$, and the running measure, F (1, 22) = 7.33, $p < .05$. Analyses of Alley 2 speeds following F trials in Goal Box 1 failed to reveal significant differences between Group 50–50 and 50–100, $F < 1$.

That Alley 2 speeds following R trials in Goal Box 1 were faster for Group 50–50 than Group 50–100 could be considered as analogous to the "crossover" effect, i.e., faster running by 50%-reinforced *Ss* that

100%-reinforced *Ss* in a straight alleyway. This finding has been explained in terms of frustration theory as reflecting the motivational effects of feedback of anticipatory frustration ($r_f - s_f$) from frustrative nonreward. It is apparent from the present data, however, that the influence of frustrative nonreward in Goal Box 2 on Alley 2 performance is exerted only on trials in which reward is present in Goal Box 1.

The postshift Alley 2 start and running speeds as a function of blocks of eight trials are presented in Fig. 2. Preshift Block 8 data for Groups 50–50 and 50–100 include only the Alley 2 start and running speeds following R trials in Goal Box 1. Analyses of variance performed on Alley 2 start and running speeds over Preshift Block 8 and Postshift Blocks 9–11 revealed that: (*a*) The Goal Box 1 reward percentage variable failed to be significant for either start speeds, F (1, 44) = 1.27, $p > .05$, or running speeds, $F < 1$; (*b*) the Goal Box 2 reward percentage variable achieved statistical significance for both start and running speeds, F (1, 44) = 16.92 and 28.74, respectively, $p < .01$; (*c*) the Blocks × Goal Box 2 Reward Percentage interaction was statistically reliable for both start and running speeds, F (3, 132) = 57.82 and 4.67, respectively, $p < .01$; (*d*) the blocks variable was significant for both start and running speeds, F (3, 132) = 57.82 and 102.48, respectively, $p < .01$; and (*d*) none of the remaining interactions achieved statistical reliability (all ps $> .05$).

The postshift data in Fig. 2 clearly indicate the typical partial reinforcement effect, i.e., greater resistance to extinction by *Ss* who previously received partial reinforcement in Goal Box 2. Failure of partial reinforcement in Goal Box 1 to facilitate resistance to extinction in Alley 2 deserves mentioning. It has been found (Brown & Logan, 1965) that when discrimination training is given with 100% reinforcement to a positive stimulus (S^+) and 50% reinforcement to a negative stimulus (S^-), resistance to extinction to S^+ is greater than for 100% reinforcement alone. To a degree, it might have been expected that partial reinforcement in Alley 1 might show a similar generalization to Alley 2 resulting in greater resistance to extinction in Alley 2. It is possible, however, that the failure of a generalized partial reinforcement effect could be attributed to the distinct physical difference between the two alleyways, i.e., a shorter black alley with a rubber floor versus a longer white alley with a hardwood floor.

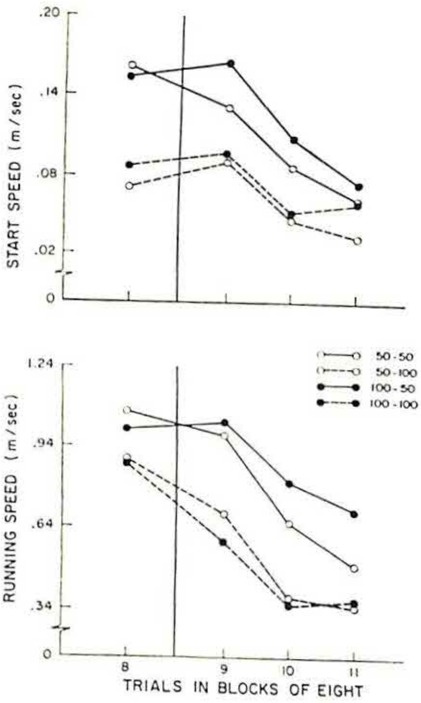


FIG. 2. Mean Alley 2 start and running speeds in blocks of eight trials during the postshift phase. (Block 8 represents the last trial block of the preshift phase.)

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

BROWN, R. T., & LOGAN, F. A. Generalized partial reinforcement effect. *Journal of Comparative and Physiological Psychology*, 1965, **60**, 64–69.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *Journal of Experimental Psychology*, 1959, **57**, 57–64.

KRIPPNER, R. A., ENDSLEY, R. C., & TACKER, R. S. Magnitude of G_2 reward and the frustration effect in a between-subjects design. *Psychonomic Science*, 1967, **9**, 399–400.

LEVY, N. & SEWARD, J. P. Frustration and homogeneity of rewards in the double runway. *Journal of Experimental Psychology*, 1969, **81**, 460–463.

SGRO, J. A., GLOTFELTY, R. A., & PODLESNI, J. A. Contrast effects and delay of reward in the double alleyway. *Psychonomic Science*, 1969, **16**, 29–31.

(Received March 13, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 232–234

EFFECTS OF MEMORY LOAD ON REACTION TIME[1]

C. F. DARLEY, R. L. KLATZKY, AND R. C. ATKINSON[2]
Stanford University

In one condition (no-cue) of a memory search task, each element of the memory set had an equal probability of matching the test stimulus on positive trials. In the other two conditions (auditory cue and visual cue), *S* was told prior to the test which element of the memory set (the "high-probability" item) would be tested. For the no-cue condition, functions relating reaction time (RT) to memory-set size were linear and increasing with a slope of approximately 38 msec. For the auditory-cue and visual-cue conditions, RT functions were essentially flat, indicating thet *S* compared the test stimulus with only the high-probability item and that memory load did not influence RT.

Sternberg (1966, 1969) has made use of a search paradigm to study the retrieval of information stored in short-term memory (STM). In this paradigm, *S* is presented a set of items (memory set) on each trial of the experiment; *S* then views a test stimulus which may or may not be a member of the memory set. The *S* is required to make a positive response if the test stimulus is a member of the memory set, and a negative response otherwise. The principal measure of performance is *S*'s reaction time (RT), defined as the interval between presentation of the test stimulus and the response. The size of the memory set is varied from trial to trial, and therefore it is possible to plot functions relating RT to memory-set size. These functions are typically found to be linear and increasing with set size. Sternberg (1969) theorizes that this search task involves four stages: (*a*) encoding of the test stimulus, (*b*) comparison of the test stimulus with members of the memory set, (*c*) choice of a positive or negative response, and (*d*) execution of the response. The intercept of the RT function depends upon the time taken to complete Stages *a*, *c*, and *d*, since these components are independent of memory-set size. The slope of the function represents the time necessary to compare the test stimulus with each member of the memory set and reflects Stage *b*.

In the typical memory-scanning paradigm, each item in the memory set has an equal probability of being matched by the test stimulus on positive trials (i.e., a probability of $1/s$ for a set size of s). Klatzky and Smith (1972) examined the effects of memory-set size and test-stimulus probability on RT. For some memory sets, *S* was told that a specific element was a high-probability item, indicating that on positive trials that element would be more likely to be matched by the test stimulus than would the other elements in the set. Klatzky and Smith found that the functions relating RT to memory-set size were linear and increasing for all types of trials. There was a significant difference in mean RTs between positive trials where a high-probability item had been designated and positive trials in which all items had an equal probability of being matched by the test stimulus. This difference was accounted for by a difference in the intercepts of the RT functions. No significant slope difference was found. Klatzky and Smith concluded that the probability information affects the encoding or response decision stages of the retrieval process. Because of the lack of a slope effect, there appears to be no biasing of the comparison stage when the probability of testing a specific element of the set is varied.

Klatzky and Smith's (1972) results are surprising; it would seem that RT on trials in which a high-probability item was designated could be reduced if *S* used the probability information to modify the comparison process. In the present experiment, an attempt was made to devise conditions such that modification of the comparison stage would lead to optimal retrieval. On a portion of the trials, *S* knew exactly which letter in the memory set would match the test stimulus. In other words, on positive trials the probability of that particular letter being matched by the test stimulus was 1.0.

The experiment involved three types of trials: no-cue, auditory-cue, and visual-cue trials. For all trials memory sets of 1, 2, 3, 4, or 5 letters were used. On no-cue trials, memory sets were presented visually and all letters in the set had an equal probability of appearing on the test. Auditory-cue trials were similar to no-cue trials in that the entire memory set was presented visually, but *E* also spoke aloud one of the members of the memory set; *S* was instructed that the spoken letter would be the only member of the memory set that could match the test stimulus, if there was a match. On visual-cue trials, the complete memory set was read aloud by *E* while one member of the set was shown visually; *S* was told that the letter presented visually was the only one that could match the test stimulus, if there was a match. The memory-set element which has a probability of 1.0 of matching the test stimulus on positive trials in the auditory-cue and visual-cue conditions will be referred to as the "high-probability" item. In order to ensure that the complete set of letters was maintained in memory throughout the trial, *S* was required to repeat aloud the memory set at the end of the trial.

It was expected that the high-probability information would have one of two effects on the comparison stage of the retrieval process. First, as in the Klatzky and Smith (1972) study, the comparison stage might be uninfluenced by S's expectancies, with the test stimulus being compared with each of the memory-set elements. Second, the comparison stage might be modified, perhaps to the extent that the test stimulus would only be compared to the high-probability item. In this case, the remaining members of the memory set would have to be held in memory while the response was made, so that they could be retrieved after completion of the response. If these additional memory-set members are not involved in the comparison process, then it is of interest to determine their effect upon RT to the high-probability item.

---

[1] This research was sponsored by National Science Foundation Grant NSF GJ-443X2.

[2] Requests for reprints should be sent to R. C. Atkinson, Department of Psychology, Stanford University, Stanford, California 94305.

Method.—The Ss were seven female students at Stanford University. Each S was paid $11.00 for participating in a training session of 1½ hr. and four test sessions of 1 hr. each.

The apparatus consisted of a two-field tachistoscope and exposure box. The visual area exposed to S measured 18 × 34 cm., and the viewing distance was approximately 60 cm. Between stimulus exposures, a dot marked the center of the viewing area. On a table to the right of S three telegraph keys were placed in an arc, separated by a distance of 2 cm. The S could comfortably rest the right arm on the table and depress the keys with the forefinger. Four of the Ss were randomly chosen to depress the key on the right for a positive response and the key on the left for a negative response; for the other Ss the conditions were reversed. The S was instructed to depress the center key until ready to execute a response, and a light on the control panel enabled E to monitor S and ensure that this instruction was followed. The S held a button in the left hand which controlled the onset of the test stimulus.

Memory sets were constructed of sizes ranging from 1 to 5 letters. The letters used to generate the sets consisted of all members of the alphabet except vowels and the letter Y. The letters were randomly assigned to memory sets and no letter was repeated within a memory set. Memory sets used in the no-cue and auditory-cue conditions were typed with an IBM Executive Registry typewriter on 5 × 8 in. white cards. The center of each set was located in the middle of S's visual area and successive letters were separated by a single space. In the visual-cue condition only 1 letter was presented visually, so each of the 20 letters used in the experiment was typed in the center of a 5 × 8 in. white card. These 20 cards were used for the memory-set presentation in the visual-cue condition and also for test-stimulus presentation in all conditions.

All Ss participated in five sessions of 132 trials each. Within an experimental session the trials were blocked according to condition, with each half-session (66 trials) containing one block each (22 trials) of the no-cue, auditory-cue, and visual-cue conditions. When the three blocks of the first half-session had been completed, the same ordering of blocks was repeated for the second half-session. The order of the blocks was arranged in a Latin square design so that each condition occurred approximately equally often in the first, second, and third portions of the half-sessions. Each memory set was presented in all conditions across Ss, with each S seeing a particular set only once. For each condition within a session, there were equal numbers of positive and negative trials for all memory-set sizes and the different memory-set sizes (1 to 5) occurred 6, 8, 12, 8, and 10 times, respectively. Memory-set sizes and positive and negative trials were randomly mixed for each condition within a session. For every memory set, one member (the high-probability item) was selected as that letter which would be either read aloud in the auditory-cue condition or shown visually in the visual-cue condition; for each memory-set size each position was selected equally often.

A trial involved the following sequence of events: (*a*) The S depressed the center response key and held it down. (*b*) The E presented the memory set. When S was sure the set was in memory she said "OK." (*c*) The E removed the card already in the tachistoscope, inserted the test card, and said "OK." (*d*) The S pushed the button held in the left hand, and after a 200 msec. delay the test stimulus was exposed for 800 msec. The onset of the test stimulus exposure coincided with the onset of a latency counter. (*e*) Using the right hand, S made the appropriate response by releasing the center key and depressing the key to the right or left, which stopped the latency counter. (*f*) Immediately after responding, S verbally reported the letters in the memory set. (*g*) The S was informed whenever an error occurred on either the yes–no response or in reporting the memory set.

Results and discussion.—Only data from the final four sessions are included in the analysis, since the first session was regarded as practice. Also, RTs for incorrect responses were excluded from the analysis. Figure 1 presents mean RT as a function of memory-set size (s) for the three conditions; results for both positive and negative responses are included. Best-fitting straight lines obtained by the least-squares method are shown for the combined data from positive and negative responses. The slopes of the positive and negative functions are not reported here, but were essentially equal in all three conditions. No differences were found between intercepts for positive and negative responses in the no-cue condition, $t(6) = 1.45$, $p > .05$; however, there were intercept differences for positive and negative responses in the auditory-cue and visual-cue conditions, $t(6) = 3.79$, $p < .005$, and $t(6) = 3.53$, $p < .01$, respectively. Error rates for the RT response are included in Fig. 1 for each memory-set size; the overall error rate was 3.2% in the no-cue condition, .5% in the auditory-cue condition, and .4% in the visual-cue condition.

The RT function for the no-cue condition is increasing and approximately linear. The slope of 37.7 msec. is similar to those obtained in other memory-scanning experiments (e.g., Klatzky, Juola, & Atkinson, 1971; Sternberg, 1966). On the other hand, the slopes for the auditory-cue and visual-cue conditions are essentially flat. The results for the

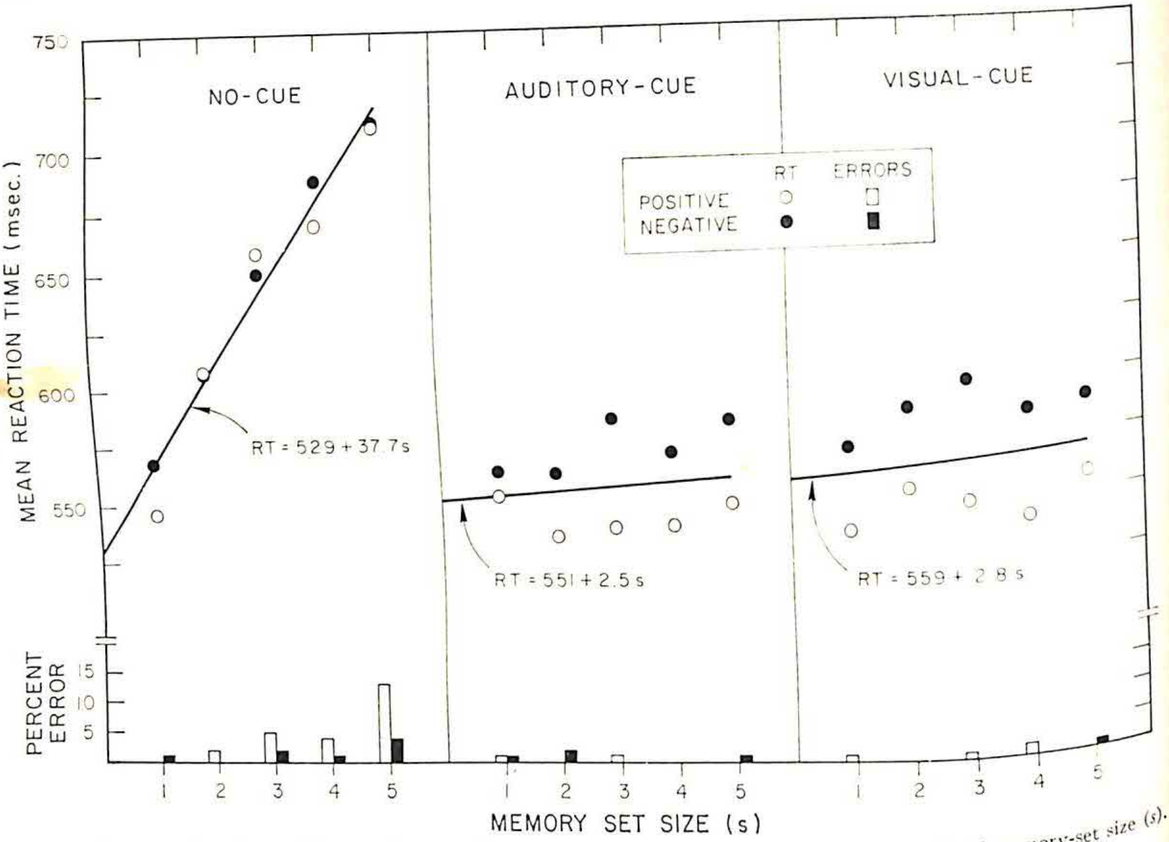


FIG. 1. Mean reaction times (RT) and error percentages for positive and negative responses as functions of memory-set size (*s*).

auditory-cue and visual-cue conditions differ from those of Klatzky and Smith (1972); this difference may be attributable to the fact that in the present study, *S*s were certain which memory-set elements would match the test stimulus, whereas in the Klatzky and Smith study it was possible for an element other than the high-probability one to match the test stimulus.

It is evident that in the auditory-cue and visual-cue conditions the memory load (i.e., the memory-set elements other than the high-probability item) had no effect on either positive or negative RTs. Clearly *S*s were not simply disregarding the memory load, since recall of the load was nearly perfect on all trials. On almost all trials *S*s recalled the memory set in correct serial order, even though they were not required to preserve the order in their recall.

There are at least three explanations for the fact that memory load has no effect on RT in the auditory-cue and visual-cue conditions. First, holding a load in STM may not interfere with the processes taking place during the RT phase of the task. Second, the maximum memory load used in the present experiment may have been too small, whereas a larger load may have shown an effect. Third, the memory load may be held in some memory system other than the one involved in making the comparison on which the RT is based.

It appears that on most trials, *S*s are performing a different task in the no-cue condition than in the auditory-cue and visual-cue conditions. However, the three conditions are virtually identical in procedure on trials where the memory set involves only one item. Because of this, it was expected that there would be no differences among RTs for the three conditions when $s = 1$; pairwise t tests were performed, and no significant differences were found between the no-cue and auditory-cue conditions, $t(6) = 1.71$, $p > .05$, the no-cue and visual-cue conditions, $t(6) = .50$, $p > .30$, or the auditory-cue and visual-cue conditions, $t(6) = 1.23$, $p > .10$.

The data presented here suggest that when *S*s are certain that a particular memory-set item will be tested, they are able to omit other members of the set in the comparison process. In fact, in the auditory-cue and visual-cue conditions *S*s performed on the RT task as if *s* were 1, regardless of the true value of *s*. It appears that the additional memory-set items had no interfering effect on the processes required for the RT task.

REFERENCES

KLATZKY, R. L., JUOLA, J. F., & ATKINSON, R. C. Test stimulus representation and experimental context effects in memory scanning. *Journal of Experimental Psychology*, 1971, **87**, 281–288.

KLATZKY, R. L., & SMITH, E. E. Stimulus expectancy and retrieval from short-term memory. *Journal of Experimental Psychology*, 1972, **94**, 101–107.

STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

STERNBERG, S. Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist*, 196 **57**, 421–457.

(Received March 22, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 235–236

RECOGNITION OF A CATEGORIZED LIST[1]

JOHN P. McLAUGHLIN[2] AND D. J. HERRMANN[3]
University of Delaware

Failures to find an effect of taxonomic organization on recognition may have been due to low degrees of list learning in those studies. The *S*s learned lists containing 15 instances of each of four categories to 95% recall. Double-word displays (DWDs) were then drawn from the lists so that DWDs contained words from one category (1C) or two (2C). Recognition latency was shorter for 1C DWDs than for 2C DWDs. Faster reading of 1C DWDs or more complicated memory search processes for 2C DWDs are alternative explanations for the effect, but category organization clearly affected the accessibility of information in memory.

Many studies have failed to find effects of organizations developed during acquisition on subsequent recognition. Those effects that have been found were small enough to prompt a recent reviewer to assume that recognition is virtually unaffected by acquisition organization (McCormack, 1972). Mandler, Pearlstone, and Koopmans (1969) have suggested that these negative results may be due to a low level of original learning. Organization is generally assumed to increase with degree of learning, but studies on recognition almost always have employed a limited amount of study.

Consequently, the purpose of the present study was to examine the effects of acquisition organization on recognition under intensive study conditions. Since such conditions preclude the use of an accuracy measure, recognition latency was chosen as the dependent variable. After acquisition, the effects of organization on recognition latency were studied by varying the representation of organizational units in test items. For example, after learning words grouped into taxonomic categories (ANIMALS, COUNTRIES), the organization of test items was manipulated by varying the number of categories represented in a two-word test item. Thus, *S*s could respond to pairs such as LION and TIGER or LION and FRANCE. If organization in acquisition affected recognition, then a pair representing a single category might be expected to be recognized more rapidly than two-category items.

Method.—Six male and six female undergraduates from the University of Delaware participated in the experiment to satisfy a course requirement. All *S*s had normal or corrected vision.

Stimuli were presented in a Scientific Prototype three-channel tachistoscope (Model GB). A white prepost exposure field was illuminated at 6.4 ftc. measured at the eye. The fixation field, a horizontal white strip, was 15.3 ftc. Probe words were presented at 17.7 ftc. on 9-in.-long, $\frac{9}{16}$-in.-wide posterboard strips. The probe words were centered, one above the other, and separated by a 1°10′ vertical visual angle. Each word subtended a 0°11′ visual angle on the vertical and an average 1°18′ on the horizontal.

Microswitch response buttons were provided to the left and right of *S*'s seat for yes (positive) and no (negative) responses. The response buttons were wired in series to a Standard Electric chronoscope, which was accurate to a hundredth of a second. Each *S* was run individually in two sessions. At the beginning of the first session, *S*s generated 15 examples for each of four categories, ANIMALS, PARTS OF THE BODY, COUNTRIES, and ARTICLES OF CLOTHING. The reason for having *S* generate this list was to provide materials for memorization which were more likely to be representative of his own taxonomic organization in long-term memory. The *S* then studied the words until he could recall at least 95% of the list. Study periods were 2 min. and recall periods were 3 min. Eight *S*s achieved 95% recall in one trial and four *S*s in two trials.

Session 2 began approximately 24 hr. later for six *S*s, 48 hr. for four *S*s, and 72 hr. for two *S*s. The *S* first restudied the list until he again reached the 95% recall criterion. All *S*s achieved this criterion in one trial. In the recognition test, which began immediately afterwards, *S* was presented with 60 double-word displays (DWDs). Half of the DWDs contained 2 words from one of the list categories and half had 2 words from two list categories.

Probe words for negative displays were selected to correspond to the taxonomic category and subcategory used on *S*'s list. For example, if *S* had only given European and Asian nations as examples of countries, negative probe words were also European and Asian nations but never South American nations. These negative probe words were drawn from the Cohen, Bousfield, and Whitmarsh (1957) norms and other sources. The other words not listed in the norms were clear examples of the experimental categories. Finally, if a DWD was composed of only one word from the list and one not from the list, *S* responded by saying "pass," a neutral response. The neutral response was necessary for DWDs to insure that *S* considered both words in a positive or negative DWD before responding.

DWD presentations began with a verbal "ready" signal from *E*, followed by a 600-msec. exposure of the fixation field, and then an approximately 4- to 6-sec. exposure of the probe display. In the presentation order of DWDs, response type (positive, negative, and neutral) was randomized in blocks of 3 DWDs. Equal numbers of one- and two-category probes as well as the specific categories used were randomly assigned to positive, negative, and neutral probes in successive blocks of 18 DWDs. Since there were 60 DWDs in total, this procedure left 6 DWDs remaining in which all conditions of category origin and response type were represented. Each category was used no less than twice and no more than three times for one-category DWDs. Each of the six possible two-category combinations were used at least once and no more than twice for two-category DWDs. A different probe presentation order was developed for each pair of *S*s. Within a pair, one *S* observed the reverse order of conditions used for the other *S*.

---

[1] These data were presented at the meeting of the Eastern Psychological Association, New York, April 1971.

[2] Requests for reprints should be sent to John P. McLaughlin, Department of Psychology, University of Delaware, Newark, Delaware 19711.

[3] The second author was a National Aeronautics and Space Administration Fellow during the conduct of this research.

Results.—The mean reaction time (RT) in seconds of correct positive responses was 1.89 for DWDs containing words from one category and 2.09 for DWDs containing words from two categories. Correct negative RT was 2.44 for one-category DWDs and was 2.68 for two-category DWDs.

It is clear that recognition took longer when DWDs originated from two different categories than when DWDs originated from a single category. Negative latency was greater than positive latency, and there was no interaction between response type and the number of categories in a DWD. Statistical analysis confirmed these observations. Recognition RT to DWDs from two categories was significantly longer than to DWDs from one category, $F\ (1, 11) = 19.0$, $p < .001$. More time was taken for negative than for positive responses, $F\ (1, 11) = 42.8$, $p < .001$. No interaction between response type and the number of categories in a DWD was observed, $F < 1.0$. The error rate was 4%.

Discussion.—The major finding of this research was that probes representing one category were correctly identified faster than probes representing two categories. Clearly the categorical organization of probes affected their recognition. There are two possible reasons for this finding. It may be that two words from one category were read faster than items from two categories. On the other hand, the organization used in acquisition may have affected a memory process at the time of testing.

Reading processes may have been affected in one of two ways. Collins and Quillian (1969) have suggested that the lexicon is organized by categories and that the comprehension of a word depends on the comparison of input with representations stored in the lexicon. If time is required to shift search from one lexical unit to another, the reading of two words stored in the same lexical unit requires less time than the reading of two words stored in different lexical units. The alternative explanation is that the reading of the first word in a one-category DWD may have elicited implicit associative responses that include other category members. Such implicit associative responses may have prepared the second word in a one-category DWD for faster reading than its counterpart in a two-category DWD.

The data, however, may not be a product of early reading processes, but rather a product of a later memory process prior to the recognition decision. When words are learned together, it may be that the comparisons in memory are more rapid when the probes represent those words learned together than when the probes do not. Neither of these explanations can be rejected in this experiment, however, since words that were learned together also belonged to the same taxonomic category.

Both of these interpretations are consistent with Tulving and Thomson's (1971) conclusion that recognition requires a retrieval process. Tulving and Thomson distinguished between an item's availability, i.e., the presence or absence of an item in storage, and its accessibility, i.e., the ease with which an item is found in storage. They assumed that retrieval operations may be inferred when the accessibility of a test item is affected by the presentation of other information held in storage. Clearly categories affected the accessibility of information in recognition tests and therefore indicate a retrieval process as defined by Tulving and Thomson.

REFERENCES

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. *Cultural norms for verbal items in 43 categories.* (Tech. Rep. No. 22) Storrs: University of Connecticut, 1957.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 240–247.

Mandler, G., Pearlstone, Z., & Koopmans, H. S. Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 410–423.

McCormack, P. D. Recognition memory: How complex a retrieval system? *Canadian Journal of Psychology*, 1972, **26**, 19–41.

Tulving, E., & Thomson, D. M. Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 1971, **87**, 116–124.

(Received November 17, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 1, 237–238

DECAY OF ACOUSTIC PROACTIVE FACILITATION[1]

LAIRD S. CERMAK[2] AND JAMES B. SAMPSON
Tufts University

Word triads presented within the context of the short-term memory (STM) distractor procedure were either completely or partially repeated on immediately adjacent trials or on trials separated by one or two unrelated triads. A complete repetition consisted of presenting exactly the same material on two separate trials, while a partial repetition consisted of presenting homonym triads (an acoustic repetition) on separate trials. While retrieval was facilitated for complete repetitions regardless of the number of intervening trials, the facilitation effect decreased across trials for the acoustic repetitions. This was interpreted as evidence that differential rates of decay exist for at least the two levels of encoding employed in this experiment.

In a recent series of short-term memory (STM) distractor type experiments, Cermak (1969, 1970) investigated the effect of a complete, or a partial, repetition of isolated verbal material on recall performance. A complete repetition was defined as the presentation of the same stimulus material on more than one trial, while a partial repetition involved presenting either identically sounding material (homonyms) or material having similar meaning (synonyms) on two separate trials.

The results of these studies demonstrated that both complete and acoustic (homonym) repetitions increased recall probability of the "repeated" items when they were presented on adjacent trials, but semantic (synonym) repetitions did not have this same proactive facilitation (PF) effect. A difference between complete and acoustic repetitions appeared when several unrelated trials intervened between the two critical repetition triads. The acoustic facilitation effect disappeared when the repetitions were separated by four unrelated trials, but the complete facilitation effect remained.

Unfortunately neither of the previous experiments (Cermak, 1969, 1970) provided a direct comparison of the decay of PF due to complete versus acoustic repetition, since different materials were used in the two studies, and the intermediate positions were never tested for either complete or acoustic repetition effects. Rate of decay of PF is important since it can be thought of as a reflection of the rate of decay of the various levels on which the initial verbal material was encoded. A comparison of rates of decay of different levels of encoding might help to determine which of two current theories of encoding is correct: Bregman's (1968) contention that the rate of decay of various levels of encoding is essentially equivalent or Tulving's (1970) proposal that the more sophisticated the level of encoding the more permanent it will be. The present experiment was designed to test these two hypotheses and to trace the rate of decay of acoustic proactive facilitation.

Method.—The *S*s were 216 introductory psychology students at Tufts University who chose to participate in the experiment as partial fulfillment of a course requirement. The *S*s were randomly assigned in order of their appearance to one of the nine groups, with 24 in each group.

A modified version of the Brown-Peterson distractor task was used. In this procedure, a word triad was projected on a screen and *S* repeated the words once; *S* was then distracted from rehearsal until a signal for recall was presented. The distractor task consisted of counting backward by three's from a number flashed on a screen in time with a metronome beating at one beat per second. Additional trials followed consisting of different word triads but using the same procedure.

The stimuli were all common words for which a homonym existed. Six such word triads, and six appropriate homonym triads, were used so that each triad could appear in each of the six possible positions in the experimental series.

The experiment consisted of four trials for three of the groups, five trials for three groups, and six trials for three groups. In all cases the material presented on Trial 3 bore one of the following relationships to the final trial for each group: Groups R—direct repetition; Groups H—homonyms; Groups C—control, no relationship. Trials intervening between Trial 3 and the final trial bore no relationship to either of these trials. In order to compare the performance of the groups on their final trial, all groups received the same stimulus material on that trial. The different relationships were established by presenting different materials on Trial 3. The R groups received the same material on both Trial 3 and the final trial, the H groups received homonyms of the final trial on Trial 3, and the C groups received words unrelated to the final trial on Trial 3. Any differences in recall on the final trial could thus be attributed to the effects of the relationships that had been produced previously rather than the difficulty level of the final trial. As stated previously, all of the slides appeared in all six positions and bore all three relationships equally.

---

[1] This work was supported in part by a Tufts University Faculty Grant BD402U to the senior author. Appreciation is extended to Janet Polivy, Carol Spadaro, and Fay Yanofsky for their help in the collection of data and preparation of this manuscript.

[2] Requests for reprints should be sent to Laird S. Cermak, who is now at the Psychology Service, Boston Veterans Administration Hospital, 150 South Huntington Avenue, Boston, Massachusetts 02130.

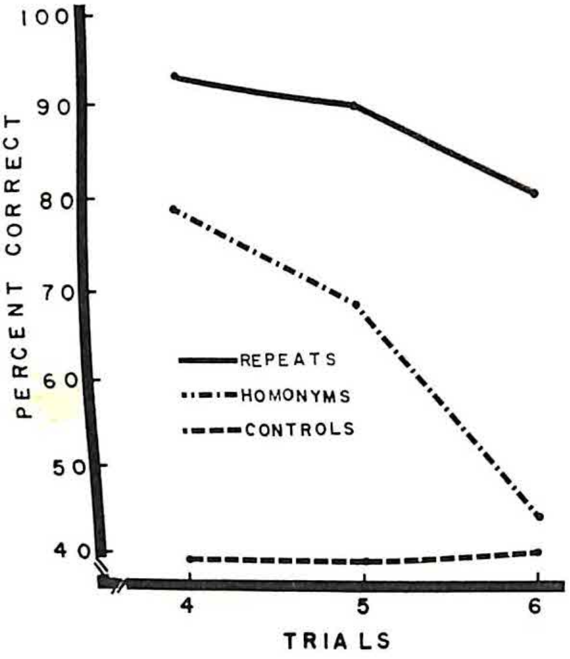


FIG. 1. Percentage of correct responses for completely repeated, acoustically repeated (homonyms), and control triads in Trial Positions 4, 5, and 6. (In all cases the related triad was presented on Trial 3.)

The *S* was seated facing a screen upon which the material was projected by means of a Kodak Carousel 800 projector. The projector was timed by a Gerbrands tape timer. The only other apparatus in the room was a metronome used to pace *S*'s distractor task. After the instructions had been given, *S*s practiced on the counting task until approximation to the metronome could be achieved. The timing and sequence of events were initiated by the tape timer for each trial: asterisk (ready signal), 2 sec., triad presentation, 2 sec., retention interval, 15 sec., recall interval, 10 sec., rest interval, 6 sec. The number of correctly recalled words was recorded for each trial.

Results.—The results for the critical trials (Trials 4, 5, and 6) are shown in Fig. 1. A two-way analysis of variance resulted in significant group, F (2, 207) = 49.63, and delay, F (2, 207) = 5.18, both ps $< .01$, differences as well as a significant interaction, F (4, 207) = 2.48, $p < .05$. Tukey multiple-comparison tests revealed that the difference between R groups and C groups was significant on each trial ($p < .01$). The difference between the H groups and C groups was significant on Trial 4 ($q = 6.81$, $p < .01$) and Trial 5 ($q = 4.89$, $p < .01$), but not on Trial 6. The difference between the R groups and H groups was significant on Trial 5 ($q = 3.55$, $p < .05$) and on Trial 6 ($q = 6.81$, $p < .01$) but not on Trial 4.

Discussion.—Proactive facilitation produced by acoustic repetition decays rapidly across trials; however, when this effect is produced by a complete repetition, it does not lose its potential to facilitate recall even after several intervening trials. Apparently something more than just the sound of a word is being encoded during the presentation interval, and it is this something extra that maintains the complete repetition's superiority in retrieval probability. Furthermore, this additional encoding has a slower decay rate than does the acoustic level of encoding (or else it begins to decay later) since otherwise its consistent ability to facilitate retrieval even after several intervening trials would not occur. This finding favors Tulving's (1970) theory of differential decay rates since the acoustic PF effect disappears more rapidly than the same effect for the complete repetition.

The advantage of the complete repetition over the acoustic repetition could come from one of two sources: First, a complete repetition involves a visual repetition, i.e., the same letter complex. Second, a complete repetition involves a semantic repetition; i.e., the same exact meaning occurs twice. While a complete repetition also involves an acoustic repetition, this does not give the complete repetition an advantage over the acoustic repetition provided by the homonyms. Thus, either the visual or the semantic level of encoding is maintaining the recall superiority of the complete repetition.

Since visual encoding has already been shown to have an extremely rapid decay rate (Sperling, 1960; Turvey, 1967), it becomes necessary to conclude that the complete repetition's PF effect is due to its being a semantic repetition. However, this conclusion immediately evokes the question of why no PF effect was found for synonym triads presented on adjacent trials (Cermak, 1970). The explanation might be that the level of semantic encoding that can be achieved within the confines of the distractor task is extremely specific and individualistic. A synonym may, or may not, be a semantic repetition to *S* despite the fact that *E* views it as a semantic repetition. However, when a complete repetition does occur, it is seen as meaning the same as a preceding triad, i.e., a semantic repetition, regardless of how *S* initially encoded it. Thus, retrieval probability for complete repetitions is always facilitated, whereas the probability that synonyms will have the same effect is quite low.

REFERENCES

BREGMAN, A. Forgetting curves with semantic, phonetic, graphic, and contiguity cues. *Journal of Experimental Psychology*, 1968, **78**, 539–546.
CERMAK, L. S. Repetition and encoding in short term memory. *Journal of Experimental Psychology*, 1969, **82**, 321–325.
CERMAK, L. S. Proactive facilitation in short term memory. *Journal of Experimental Psychology*, 1970, **85**, 305–310.
SPERLING, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, **74** (11, Whole No. 498).
TULVING, E. Short- and long-term memory: Different retrieval mechanisms. In K. H. Pribram & D. Broadbent (Eds.), *Biolog of memory*. New York: Academic Press, 1970.
TURVEY, M. T. Repetition and the preperceptual information store. *Journal of Experimental Psychology*, 1967, **74**, 289–29

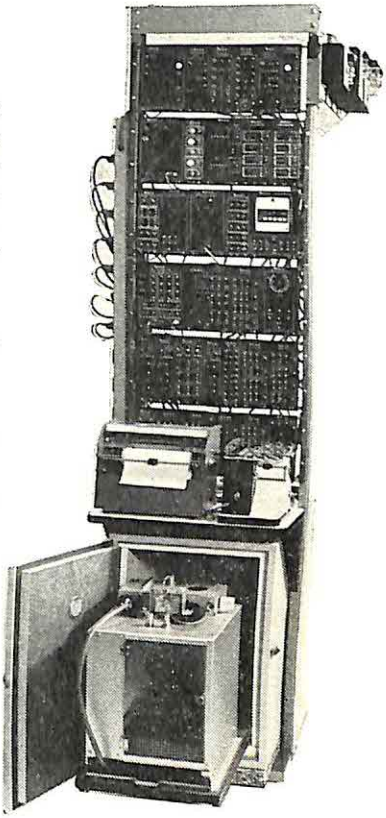
(Received December 13, 1971)

"Why should I buy Gerbrands relay programming equipment when I have been using other electro=mechanical systems for so many years?"

Why? Because most electromechanical programming systems presently in use were originally designed many years ago, and with most manufacturers concentrating on solid state and computer systems, few innovations have been added to relay programming equipment.

Until now.

Our relay programming line is completely new. It is easy to use (no engineering symbols or jargon), it is reliable, and it is innovative. Innovative with such new items as a Digital Interval Timer with push button control, a combination one per second/ten per second Pulse Generator with solid state switching, a Cam Stepper that incorporates two different mixed sequence steppers as well as an alternating stepper, and a rack mount Masking Noise Generator that sells for the low price of $75.00 fob Arlington, Massachusetts.



In addition, ours is an expanding line. As new technical developments become available we intend to incorporate these advances in our relay programming line wherever possible, in order to provide you with new and innovative instruments designed to make your electromechanical programming system more versatile and easier to program.

If you are thinking of purchasing a new electromechanical programming system or are looking for ways in which you can improve and expand your present system, write or phone for our latest catalog.

Ralph Gerbrands Company

8 BECK ROAD, ARLINGTON, MASSACHUSETTS 02174 (617) 648-6415

Accept this FREE 144-page guide

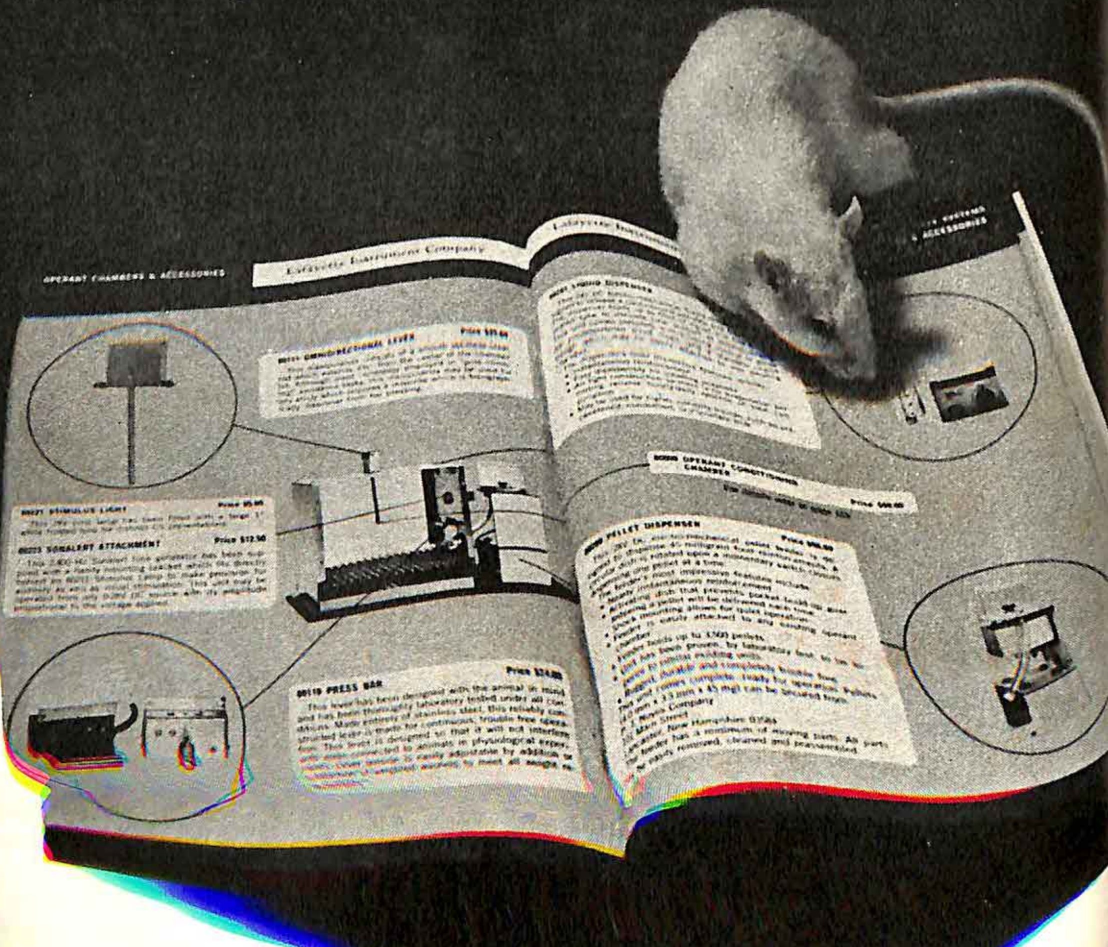
The new 1973 Lafayette Instrument Company guidebook of scientific and laboratory instruments for the life sciences.

On page 106, for instance, you'll see how to set up a custom student operant conditioning chamber. And that's just one page.

Each page is filled with facts, prices, and information about our newest quality instruments. So naturally, we'd like you to get your hands on our guidebook.

Send for your personal copy today. Ask for the 1973 edition of guidebook no. 1170.

LAFAYETTE INSTRUMENT COMPANY
P.O. Box 1279/Lafayette, Indiana 47902/ 317-423-1505


OPERANT CHAMBERS & ACCESSORIES
Lafayette Instrument Company
STIMULUS LIGHT
SONALERT ATTACHMENT
PRESS BAR
OPERANT CONDITIONING CHAMBER
PELLET DISPENSER

Journal of
Experimental Psychology

VOL. 96, NO. 2 DECEMBER 1972

HIGH-SPEED MEMORY SCANNING: STABILITY AND GENERALITY[1]

P. J. SIMPSON[2]

University of Sussex, Brighton, England

Three experiments were designed to investigate performance in S. Sternberg's recognition memory task when (*a*) *S*s were given extended training with nonintersecting positive sets containing items that were never associated with the negative response, (*b*) the positive and negative sets contained items drawn from different categories of stimuli, and (*c*) the size of the positive set was increased up to the rehearsal span for letters. Mean reaction time was a linear function of $\log_2$ (positive set size) in each case, and thus the results fail to replicate Sternberg's findings. It is argued that the results are consistent with the idea that *S*s employ a binary matching procedure to select a positive set item that is visually similar to the test stimulus and then compare the selected item with the test stimulus in a matching operation underlying response selection.

Sternberg (1966) reported that the time taken to decide if a test stimulus were a member of a short list of digits held in memory was a linear function of list length; and he suggested that performance was mediated by a serial scanning operation in which the test stimulus was compared with the memory items. He found that the estimate of comparison rate obtained when the lists were held in memory for a few seconds was equal to that obtained when *S*s were trained with fixed memory sets for 180 trials. Sternberg argued that the process underlying performance when the digit lists were held in memory for a short period also mediated performance when the digits were held in memory for a number of minutes.

However, when Sternberg's (1966) *S*s were trained with fixed memory sets, each digit was associated with both the positive and negative response during the course of the experiment. Thus the training *S*s received with digits making up the positive set on one condition could interfere with training on the remaining experimental conditions and limit the performance achieved. Evidence in favor of this view is provided by two studies in which the stimuli making up the positive sets were never associated with the negative response during the course of the experiment. Briggs and Blaha (1969) found that when *S*s were trained to associate sets of random shapes with the positive response and were presented with a test stimulus made up of a single random shape, performance after 128 training trials was associated with a

---

[1] This paper was written while the author was employed on a Science Research Council grant for work on "Visual encoding mechanisms." The computing facilities were provided by the Science Research Council. The author wishes to thank N. S. Sutherland and A. D. Baddeley for advice in the preparation of this paper.

[2] Requests for reprints should be sent to P. J. Simpson, who is now at the Department of Engineering, University of Cambridge, Cambridge, England.

© 1972 by the American Psychological Association, Inc.

TABLE 1

Composition of the Positive and Negative Sets in Experiment I

Set	Size of positive set		
	1	2	4
Positive	R N D	D,N B,Z K,T	B,K,T,Z D,K,R,T B,N,R,Z
Negative	C,F,J M,P,S G,U,H	C,F,J M,P,S G,U	C,F,J M,P,S

slope estimate comparable to that obtained by Sternberg. But after 256 trials the slope estimate was half of that obtained by Sternberg. In fact, performance after 256 trials was not well described by a linear function. The mean reaction time obtained with a one-item memory set was equal to that obtained with a two-item memory set. The associated error rates were below 1%.

Ross (1970) gave *S*s extended training in a single-character classification task with fixed memory sets made up of letters and digits. Ross found that the data were well described by a linear function when mean reaction time was plotted against $\log_2$ (positive set size). The slope estimate was 26.7 msec. per log unit after 120 trials, 25.4 msec. per log unit after 400 trials, and 22.3 msec. per log unit after 800 trials.

Experiment I

Experiment I was designed to investigate the effects of training with nonintersecting positive sets which were exclusively associated with the positive response using an experimental procedure directly comparable to that used by Sternberg (1966). In Exp. I, the positive and negative sets were made of letters of the alphabet so that *S*s were presented with highly familiar stimuli drawn from a single category.

Method

Subjects.—The *S*s were six right-handed undergraduates at the University of Sussex. None of the *S*s had received formal training in psychology or had taken part in a character-classification task before. The *S*s were paid at the rate of £.25 per hour.

Stimuli and apparatus.—The stimuli were uppercase consonants, and each *S* was trained with fixed positive sets containing one, two, and four items. Three different sets of items were selected and a subgroup of two *S*s was trained with each selection. The composition of the positive and negative sets is presented in Table 1. The items associated with the negative response were constant for all *S*s, and the number of negative set items varied with the size of the positive set to ensure that 10 different stimuli were presented in each condition. The test stimulus was displayed on an in-line display unit and subtended an angle of 1.6° at the viewing distance of 100 in. The experiment was run under the control of an Elliott 4130 computer. A 1-sec. warning light was displayed before the onset of the test stimulus and, following *S*'s response, the test stimulus was turned off and a red or green feedback light was displayed until the next trial was initiated. The *S* responded by operating a vertically mounted Bulgin Type M microswitch with the first finger of each hand.

Procedure.—Each *S* was run individually for three 1-hr. experimental sessions which took place on consecutive days. During each session, *S*s were presented with three conditions in which they were trained with fixed positive sets containing one, two, and four items. Each *S* was presented with the experimental conditions in a different order on the three experimental sessions, and the orders were balanced for the group as a whole on each session to control for practice and fatigue effects. The *S*s received 160 trials in each condition, and the trials were presented in blocks of 32 with an intertrial interval of 4 sec. On 50% of trials controlled over blocks of 16 trials, the test stimulus was a member of the positive set. The *S*s were instructed to associate the positive set with the right-hand response, and the experimental instructions stressed that *S*s should respond swiftly and only make an occasional error. The experimental procedure differs from that used by Sternberg (1966) in the following respects: (*a*) Items in the positive sets were exclusively associated with the positive response throughout the experiment. (*b*) Positive and negative responses were equally probable. In Sternberg's experiment with fixed positive sets the positive response was required on only 25% of trials. However, data published by Sternberg (1969) indicate that the slope estimate is not reliably affected when the relative frequency of the positive response is changed from .25 to .5. (*c*) Six of the items associated with the negative response were common to all three experimental conditions. (*d*) No direct financial incentive was offered for accurate or fast performance.

Results

The first 64 trials presented in each condition during an experimental session were regarded as practice. The results presented in Fig. 1 are based on the correct responses that occurred in the later 96 trials in each condition. The mean error rates obtained in each condition are presented in Table 2. All six sets of data points exhibited significant linear trends ($p < .001$ in all cases). Linear regression accounts for over 99.6% of the variance for each line presented except in the case of the data for positive responses on the third experimental session when linear regression accounts for 97.7% of the variance. Trend tests based on orthogonal polynomials (Edwards, 1966) were carried out on the positive set size variable separately for each response type and the three experimental sessions as a further test of the linearity of the functions. The tests revealed a significant linear component in all cases ($p < .01$), but the quadratic components were not significant. A four-way analysis of variance was carried out with *S*s, size of the positive set, response type, and experimental session as factors. All of the main effects were significant ($p < .01$), but none of the second- or third-order interactions were significant.

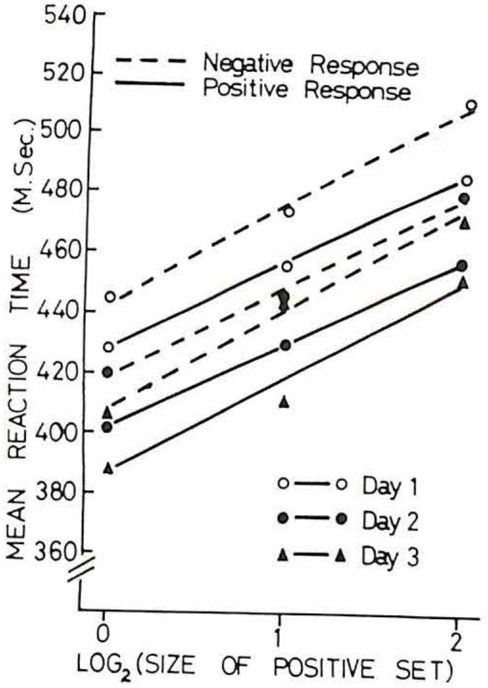


FIG. 1. Mean reaction time as a function of positive set size, response type, and practice in Exp. I.

Discussion

The results most closely parallel those obtained by Ross (1970). The data are well described by a linear function when mean reaction time is plotted against $\log_2$ (positive set size). Training over three experimental sessions affected the intercept but not the slope. The average slope was 30 msec. per log unit. Thus the results fail to replicate Briggs and Blaha's (1969) observation that after training the mean reaction time obtained with one- and two-item positive sets became equal. However, taken as a whole, these three studies suggest that when *S*s are given training comparable to that given by Sternberg (1966) with fixed memory sets that are exclusively associated with the positive response, mean reaction time is no longer a linear function of positive set size and the increase in mean reaction time that occurs when the positive set is increased from one to four items is substantially less than that obtained by Sternberg. By contrast, those studies which employ Sternberg's experimental procedure produce data that are characterized by a linear relationship between mean reaction time and positive set size and higher slope estimates. Chase and Calfee (1969) reported slope estimates of 47 and 41 msec. per item in a short-term recognition memory task comparable to that used by Sternberg (1966). Sternberg (1966, 1967) obtained slopes of 38 msec. per item, and Bracey (1969) reported slopes of 38 and 31 msec. per item. In terms of contemporary models of memory, training may affect the form in which information is stored and the method by which it is accessed. Sternberg (1967) has suggested that the slope estimates obtained in character classification

TABLE 2

PERCENTAGE OF ERROR RATES IN EXPERIMENT I

Day	Response type	Positive set size		
		1	2	4
1	Positive	3.0	1.6	3.5
	Negative	2.5	1.6	2.6
2	Positive	1.6	4.0	3.0
	Negative	2.0	.6	.3
3	Positive	2.3	3.3	3.3
	Negative	1.3	.6	1.6

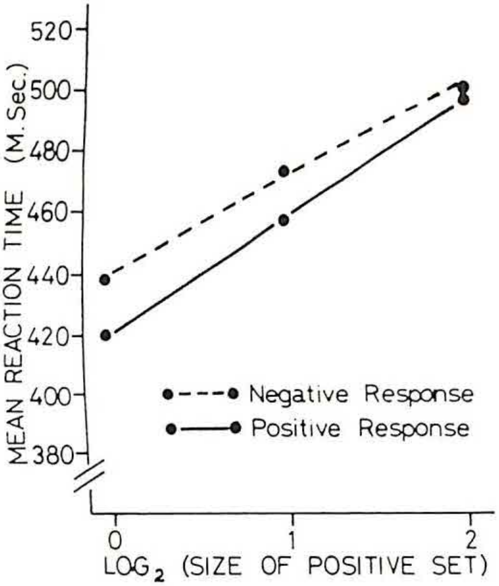


FIG. 2. Mean reaction time as a function of positive set size and response type in Exp. II.

tasks may depend on the type of representation employed by *S* to store the memory items. The findings obtained in Exp. I, together with those obtained by Briggs and Blaha, and Ross, indicate that after prolonged training on a character-classification task, retrieval of information may not be achieved by a serial matching operation.

EXPERIMENT II

This experiment investigated how performance in Sternberg's (1966) task was affected by presenting *S* with positive and negative sets containing items which are already highly differentiated as a result of *S*'s extraexperimental experience. The positive sets consisted of one, two, and four digits, and the stimuli associated with the negative response were letters. Thus *S* could base his decisions on the well-established skill of categorizing stimuli as letters and digits.

Method

Subjects.—The *Ss* were six right-handed undergraduates at the University of Sussex. None of the *Ss* had received formal training in psychology or had taken part in a character-classification task before. The *Ss* were paid at the rate of £.25 per hour.

Stimuli and apparatus.—The *Ss* were trained with fixed nonintersecting positive sets containing one, two, or four digits. An unspecified set of upper-case letters was associated with the negative response. Two different selections of digits were made to construct the positive sets, and a subgroup of three *Ss* was trained with each selection. The composition of the positive and negative sets is presented in Table 3. The digits 0 and 1 and the letters O and I were excluded. The number of letters associated with the negative response varied with the size of the positive set to ensure that 10 different stimuli were presented in each condition. The experiment was run under the control of an Elliott 4130 computer using a visual display unit to generate the stimuli. A .5-sec warning-fixation spot was presented before the onset of the test stimulus, which was also displayed for .5 sec. The *Ss* viewed the display unit at a distance of 30 in., and the test stimuli subtended a visual angle of .5° at the eye. The *Ss* responded by operating a vertically mounted microswitch with the first finger of each hand.

TABLE 3

COMPOSITION OF THE POSITIVE AND NEGATIVE SETS IN EXPERIMENT II

Set	Size of positive set		
	1	2	4
Positive	2 5	7,8 4,9	3,4,5,9 2,6,7,8
Negative	B,D,F G,K,R S,Y,Z	B,F,G K,R,S Y,Z	B,G R,S Y,Z

Procedure.—Each *S* was run individually for one experimental session which lasted approximately 1 hr. During the session, *Ss* were presented with three conditions in which they trained with fixed positive sets containing one, two, and four items. Each *S* was presented with the conditions in a different order to balance out practice and fatigue effects. The *Ss* received 192 trials, presented in six blocks of 32 trials, in each condition except in the case of the first condition received, when an additional 32 trials were presented to familiarize *Ss* with the experimental procedure. The *Ss* were told to associate the positive set with the right-hand response. The experimental instructions stressed that *Ss* should respond swiftly but with only an occasional error. No direct financial incentive was given for fast or accurate performance.

Results

The results are presented in Fig. 2. They are based on the correct responses that occurred in the last 96 trials presented in each condition. The mean error rates obtained with the one-, two-, and four-item positive sets were 0%, 2%, and 3% for positive responses and 0%, 1%, and 3% for negative responses. Linear regression accounts for over 99.7% of the variance for each line presented. Trend tests, based on orthogonal polynomials, carried out on the positive set size variable for each response type revealed significant linear components ($p < .01$), but the quadratic components were not significant. A three-way analysis of variance was carried out with *S*s, positive set size, and response type as factors. All the main effects were significant ($p < .001$), but the interaction was not significant.

Discussion

The results of Exp. II are very similar to those of Exp. I. Mean reaction time is a linear function of $\log_2$ (positive set size), and the slope estimate is 37 msec. per log unit for the positive response and 30 msec. per log unit for the negative response. The results suggest that the operations underlying performance in a character-classification task with positive sets containing up to four items are unaffected by the well-established skill which permits *S*s to categorize stimuli as letters or digits. It appears that the most efficient strategy *S*s could have employed in response to the experimental instructions was essentially the same as that used in Exp. I.

EXPERIMENT III

Sternberg (1970) has suggested that since the slope estimate obtained with his experimental procedure was unaffected by training with fixed memory sets, *S*s store the items in an active or working memory. He suggested that *S*s transfer information from long-term memory into active memory because this will presumably simplify the retrieval operation and reduce the time taken to respond. This explanation could also account for the similarity of the results obtained in Exp. I and II described above.

If *S*s in Sternberg's task (1966) make use of an active memory in which capacity is limited by the rehearsal span, then performance might be disrupted when the number of items in the memory set approaches or exceeds the rehearsal span. The data presented by Sternberg (1966) are based on experiments in which *S*s were either required to remember up to six digits for a number of seconds or up to four digits for many minutes. Since the error rate associated with performance was very low, memory sets of 4 and 6 items appear to be within the rehearsal span for active memory. Wingfield and Branca (1970) presented *S*s with memory sets containing up to 12 items in a short-term recognition memory task. They found that mean reaction time increased linearly with memory set size up to 10 items, but the error rates associated with the 8-, 10-, and 12-item memory sets were 10%, 12%, and 20%, respectively. These results suggest that when the memory set contained 8 or more items *S*s could not maintain all the items in active memory and were forgetting some of the items. When *S*s are trained with fixed memory sets, the items are stored in long-term memory. If the number of items in the memory set exceeds the rehearsal span for active memory, *S* is not able to maintain all the items in active memory. Thus *S*s could be forced to interrupt the matching operation to transfer the remaining items into active memory or match the test stimulus with items stored in long-term memory, and reaction time should increase as a result. Ross (1970) found that mean reaction time was a linear function of $\log_2$ (memory set size) with memory sets containing up to 8 items. This finding suggests that when *S*s are trained with fixed positive sets, the rehearsal span for active memory is at least 8 items. However, the error rates of 6% and 7% obtained with 4- and 8-item sets are higher than those obtained by Sternberg. Experiment III investigated performance after training with memory sets containing 1, 4, and 8 items when the items are drawn from a single category and the error rates are comparable to those obtained by Sternberg.

Method

Subjects.—The *S*s were six right-handed undergraduates at the University of Sussex. None of the *S*s had received formal training in psychology or had taken part in a character-classification task before. The *S*s were paid at the rate of £.25 per hour.

Stimuli and apparatus.—The *S*s were trained with

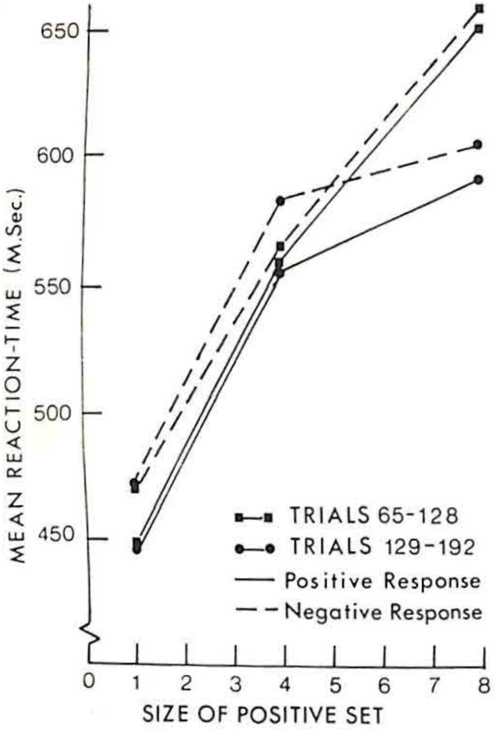


FIG. 3. Mean reaction time as a function of positive set size, response type, and number of trials in Exp. III.

fixed positive sets containing one, four, and eight uppercase letters. An unspecified set of eight letters was associated with the negative response in each condition. In the case of the one- and four-item positive sets, two different selections of letters were made and a subgroup of three *S*s was trained with each selection. The composition of the positive and negative sets is presented in Table 4. The apparatus, warning-spot and test-stimulus duration, and intertrial interval were identical to those used in Exp. II.

Procedure.—Each *S* was run individually for one experimental session which lasted approximately 1 hr. At the beginning of the session, *S*s were given 32 trials training with a positive set containing two digits and a negative set made up of letters in order to familiarize them with the experimental procedure. They were then presented with three experimental conditions in which they trained with fixed positive sets containing one, four, and eight items. Each *S* received the experimental conditions in a different order to balance out practice and fatigue effects. The *S*s received 192 trials, presented in six blocks of 32 trials, in each condition. On 50% of trials, controlled over blocks of 16, the test stimulus was a member of the positive set. The *S*s were told to associate the positive set with the right-hand response, and the experimental instructions stressed that they should respond swiftly but with only an occasional error.

Results

The results presented in Fig. 3 are based on correct responses which occurred during Trials 65–128 and Trials 129–192. The mean error rates obtained during Trials 65–128 with one-, four-, and eight-item positive sets were 2.1%, 1.0%, and 7.8% for positive responses and 1.0%, 1.5%, and 4.7% for negative responses. The corresponding error rates obtained during Trials 129–192 were 0%, 2.1%, and 2.1% for positive responses and 0%, 1.0%, and 3.1% for negative responses.

Trials 65–128.—The data obtained for positive responses exhibited a significant linear trend when reaction time is plotted against both positive set size and $\log_2$ (positive set size). However, when reaction time is plotted against $\log_2$ (positive set size), the percentage of variance accounted for by linear regression (99.6%) and the reliability of the linear trend, $F = 289$, $p < .001$, are greater than those obtained when reaction time is plotted against positive set size. Linear regression accounts for 95.8% of the variance and $F = 22.3$, $p < .025$, when reaction time is plotted against positive set size.

The data obtained for negative responses also exhibit significant linear trends when reaction time is plotted against positive set size and $\log_2$ (positive set size). In this case, the percentage of variance accounted for (99.2%) and the reliability of the linear trend, $F = 118$, $p < .01$, obtained when reaction time is plotted against positive set size are slightly greater than those obtained when mean reaction

TABLE 4

COMPOSITION OF THE POSITIVE AND NEGATIVE SETS IN EXPERIMENT III

Set	Size of positive set		
	1	4	8
Positive	K	B,G,N,Y	B,F,G,K N,P,W,Y
	P	F,K,P,Y	
Negative	C,D,H M,R,T X,Z	C,D,H M,R,T X,Z	C,D,H M,R,T X,Z

time is plotted against $\log_2$ (positive set size). Linear regression accounts for 98.6% of the variance and $F = 58.5$, $p < .01$, when reaction time is plotted against $\log_2$ (positive set size).

The mean slope obtained when reaction time is plotted against positive set size is 26.2 msec. per item. When reaction time is plotted against $\log_2$ (positive set size), the mean slope is 59.8 msec. per log unit.

Trials 129–192.—The data obtained during this block of trials only exhibit significant linear trends for both the positive responses, $F = 75.7$, $p < .01$, and negative responses, $F = 82$, $p < .01$, when mean reaction time is plotted against $\log_2$ (positive set size). Linear regression accounts for over 98.7% of the variance in each case. The mean slope was 47.4 msec. per log unit.

Discussion

Extended training on the task presented to the *S*s in Exp. III changes the form of the relationship between mean reaction time and positive set size. Training primarily affected the reaction time and error rate associated with the eight-item memory set condition. A four-way analysis of variance was carried out with *S*s, positive set size, response type, and number of trials (128 or 192) as factors. Positive set size ($p < .01$) and the Practice × Positive Set Size interaction ($p = .025$) were significant sources of variance. The results obtained for Trials 129–192 add further support to the finding that after extended training with fixed positive sets mean reaction time is a linear function of $\log_2$ (positive set size), and suggest that if *S*s make use of some form of active or working memory the rehearsal span is at least eight items. However, it should be noted that the mean slope of 47.4 msec. per log unit associated with performance during Trials 129–192 is substantially higher than the mean slope obtained in Exp. I.

GENERAL CONCLUSION

In Exp. I and III, it was found that when *S*s were trained with fixed positive sets containing items which were never associated with the negative response, mean reaction time for positive and negative responses was a linear function of $\log_2$ (positive set size). The same results were obtained in Exp. II when the positive and negative sets contained items drawn from different categories. The similarity of the results obtained for positive and negative responses suggests that even after extended training, *S*s select a response by executing an operation in which the test stimulus is matched with the positive set items and the matching operation is exhaustive. However, the relationship between mean reaction time and positive set size obtained in these experiments is not consistent with the serial matching operation proposed by Sternberg (1966). The results suggest that the decisions underlying response selection have been reorganized. Therefore, since Sternberg (1967) has argued that *S*s employ visual representations of the test stimulus and positive set items after training, the results suggest that extended training modifies the process in which *S*s match the test stimulus with visual representations of the positive set items.

The relationship between mean reaction time and positive set size obtained in these experiments could be interpreted in terms of the binary matching procedure discussed by Hick (1952) and Welford (1960). The binary matching model was originally introduced to account for performance in a choice reaction-time task in which a single stimulus was mapped onto each response. Hick proposed that his data were consistent with the view that *S*s categorize the stimulus by applying a series of subdecisions each taking approximately the same time. With the first subdecision, *S*s identify the stimulus and the response to it as lying within one half of the total possibilities, with the second subdecision as lying within one half of this half, and so on until the actual stimulus, and hence the response to it, has been located in memory. However, Welford has argued that it seems unlikely that *S*s can develop a decision procedure which will discriminate between the alternative stimuli in the ideal way envisaged by Hick. In practice, a binary decision procedure may only permit a partial categorization of the test stimulus presented for classification. Thus, although a binary matching procedure could enable *S*s to select the most likely alternative from among the stimuli held in memory, an additional operation in which the selected item is itself matched with the test stimulus may be required before a response can be generated. In the context of a recognition memory task, a binary matching procedure could take the form of a series of matching operations in which *S*s initially check the test stimulus for visual features which characterize subsets of the items in the positive set. This decision procedure would provide *S*s with partial information about the structure of the test stimulus that could be used to select a positive set item which is visually similar to the test stimulus. The positive set item selected by the binary matching procedure could then be compared with the test stimulus in the final matching operation underlying response selection. If the time taken to check for a visual feature of the test stimulus is comparable to that needed to match a positive set item with the test stimulus, then a binary procedure would offer no advantage over a serial matching operation with a two-item positive set. However, a binary procedure would reduce the number of decisions underlying performance with a four-item positive set to three, i.e., two binary decisions plus one matching operation between the test stimulus and a positive set item. With an eight-item positive set, the number of decisions could be reduced to four, i.e., three binary decisions plus one matching operation. Thus a binary model predicts that the relationship between mean reaction time and positive set size should be well characterized by a linear function when reaction time is plotted against $\log_2$ (positive set size) and is compatible with the results obtained in the present set of experiments.

REFERENCES

BRACEY, G. W. Two operations in character recognition: A partial replication. *Perception & Psychophysics*, 1969, **6**, 357–360.

BRIGGS, G. B., & BLAHA, J. Memory retrieval and central comparison times in information processing. *Journal of Experimental Psychology*, 1969, **3**, 395–402.

CHASE, W. G., & CALFEE, R. C. Modality and similarity effects in short-term recognition memory. *Journal of Experimental Psychology*, 1969, **81**, 510–514.

EDWARDS, A. L. *Experimental design in psychological research*. Holt, Rinehart & Winston, 1966.

HICK, W. E. On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 1952, **4**, 11–26.

ROSS, J. Extended practice with a single-character classification task. *Perception & Psychophysics*, 1970, **8**, 276–278.

STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

STERNBERG, S. Two operations in character recognition. Some evidence from reaction time measurements. *Perception & Psychophysics*, 1967, **2**, 45–53.

STERNBERG, S. The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 1969, **30**, 276–315.

STERNBERG, S. Memory-scanning: Mental processes revealed by reaction-time experiments. In J. S. Antrobus (Ed.), *Cognition and affect*. Boston: Little, Brown, 1970.

WELFORD, A. T. The measurements of sensory-motor performance: Survey and reapprasial of twelve years progress. *Ergonomics*, 1960, **3**, 189–230.

WINGFIELD, A., & BRANCA, A. A. Strategy in high-speed memory search. *Journal of Experimental Psychology*, 1970, **83**, 63–67.

(Received February 26, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 247–254

TESTS OF AN ALL-OR-NONE MODEL OF VERBAL MEDIATED RESPONDING[1]

KENT L. NORMAN[2] AND IRWIN P. LEVIN

University of Iowa

Choice behavior was investigated in a multiple-choice task that requires *S* to choose the alternative that "goes with" a stimulus word, where one alternative may be chosen on the basis of an experimentally induced mediator. A mathematical model was proposed as an initial attempt to understand how the mediation process proceeds over trials. The main assumption of the model is that *Ss* come to use the same mediator consistently from one trial to the next in an all-or-none manner. Experiment I used a double-function list of common nouns for the acquisition stage and included a mixed- and an unmixed-list presentation of test items. Experiment II used a three-stage chaining paradigm with stages of the type paralog-noun, noun-trigram, and paralog-trigram, in which the third stage (test list) contained either two or three response alternatives. Reliable mediation effects were obtained in each experiment, and the all-or-none assumption was supported in each experiment. The number of free parameters necessary to account for the data was determined for each experiment and the effects of the variables of interest were related to specific parameters.

In a test for mediated responding, *Ss* learn the paired associates A-B and B-C and then are given a test-only multiple-choice task with A as the stimulus and C as one of the response alternatives. If C is chosen at a greater-than-chance level, mediated responding is said to have occurred. Mediated responding is distinguished from mediated facilitation in that it is not evidenced in a learning task but in a test of associative strength. The *S* is given a choice task in which he is instructed to give the alternative that is "best paired" or "goes with" the stimulus term. The *S* receives no feedback as to the correct or incorrect response. A number of investigators have produced reliable mediated responding using a variety of lists and procedures (e.g., James & Hakes, 1965; Norman, Levin, & Williams, 1970; Peterson & Blattner, 1963; Weaver, Hopkins, & Schulz, 1968).

The typical dependent measure used to indicate mediated responding is the number of choices of a chaining alternative (the alternative that may be chosen on the basis of an experimentally learned mediator) in a fixed number of trials. Earlier investigations typically were designed to demonstrate qualitatively that mediated responding does occur. Little attention has been given to a quantitative analysis of the process by which *S* arrives at his choices. Now that mediated responding in a paired-associates chaining paradigm has been established as a reliable phenomenon, such quantitative analyses would appear to be warranted. The present approach to the problem of analysis is the application of a mathematical model capable of making detailed predictions and estimating parameters that can be tied to independent variable manipulations.

The present interpretation of the process underlying mediated responding is that when *S* is presented with Stimulus A, he may "sample" an associate to A. This associate may be the experimentally learned mediator B or some preexperimental natural language or idiosyncratic associate.

---

[1] This report is based on a thesis submitted to the Graduate College of the University of Iowa in partial fulfillment of the requirements for the master's degree by the first author under the supervision of the second author. This project was completed during the first author's tenure in the United States Public Health Service Research Predoctoral Fellowship, National Institute of Mental Health Grant MH-51134-01.

[2] Requests for reprints should be sent to Kent L. Norman, Department of Psychology, University of Iowa, Iowa City, Iowa 52240.

This implicit associate then evokes a recognition of one of the alternatives. If *S* is sampling from a pool of associates, there is a nonzero choice probability associated with each of the alternatives on a given trial. At some point, *S* may narrow the pool down to one associate and thus choose one of the alternatives consistently. Though other studies have attempted to indicate whether or not *S*s mediate in an all-or-none manner over items (i.e., use the same "rule" on all of the items in the list), the present study asks whether *S*s come to choose a mediator in an all-or-none manner for a given item.

A Markov chain with absorbing states will be used to model behavior in the choice task. Each item may be in one of three states. The first state is a random-choice state, R, in which *S* chooses Alternative C with probability $g_c < 1$. In this state, *S* samples from a pool of implicit associates. The other two states are absorbing states in which *S* consistently chooses Alternative C or consistently chooses a Nonchaining Alternative X. If the item is in the C-choice state, C, the Alternative C is given consistently, and if the item is in the X-choice state, X, an Alternative X is given consistently. The probability of leaving State R to enter one of the absorbing states on any one trial will be designated c. The probability that an item ends in State C will be designated p, and the probability that an item ends in State X will be $1 - p$. Thus, on any one trial, pc is the probability of going from State R to State C, and $(1 - p)c$ is the probability going from State R to State X. These transition probabilities are assumed not to change over trials. The probability of being in a particular state on Trial 1 will be assumed to be the same as the transition probability from State R to that state since all test-only trials are identical. The matrix of transition probabilities for the three states, the probabilities of being in each state on the first trial, and the probabilities of choosing Alternative C, $P(C)$, in each state are as follows:

State on Trial $n + 1$

$$
\begin{array}{lc}
 & \begin{array}{ccc} \quad C & \quad R & \qquad X \end{array} \qquad \begin{array}{c}\text{Start}\\ \text{Vector}\end{array} \qquad P(C) \\
\begin{array}{l}\text{State on}\\ \text{Trial } n\end{array} \quad \begin{array}{c} C \\ R \\ X \end{array} & \begin{bmatrix} 1 & 0 & 0 \\ pc & 1-c & (1-p)c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} pc \\ (1-c) \\ (1-p)c \end{bmatrix} \begin{bmatrix} 1 \\ g_c \\ 0 \end{bmatrix}
\end{array}
$$

Two experiments with widely different lists and paradigms were conducted to test the basic all-or-none assumption of the model and to determine whether or not the model provides an adequate description of the data. Several independent variables were manipulated and their effects on the model's parameters were assessed. In Exp. I, a procedure similar to that used by Norman et al. (1970) was employed. The *S*s learned a 16-item double-function list containing A-B and B-C associates. Four chaining items of the type A − C, X were presented as an unmixed test list for one group of *S*s. A mixed-list group received the four chaining items plus an additional four nonchaining items of the type A − X_1, X_2 in which neither alternative could be chosen on the basis of an experimentally induced mediator. The test list was repeated until the same alternative had been given on five consecutive trials for each item. This criterion corresponds to the assumption that the item has entered an absorbing state. In Exp. II, a three-stage chaining paradigm was used with lists similar to those introduced by Schulz, Weaver, and Ginsberg (1965). The number of alternatives per item in Stage 2 was set at two for one group and at three for another group.

METHOD

Experiment I

Materials.—The materials were the same as those used by Norman et al. (1970) and are described briefly here. The double-function list was composed of high-meaningful nouns. The list of 16 word pairs was constructed so that two out of every six of the nouns served as both stimuli and responses. Within the 16 pairs there were four instances of each of four types, which may be designated A-B, B-C, C-D, and E-F. Two variations of the list were constructed to counterbalance chaining and nonchaining items. Of the eight mixed-list test items, half were nonchaining items, in which neither of the two response alternatives could be chosen on the basis of a mediating chain. An example of a chaining item would be DANGER paired with FOOT, and FOOT with CUP in the double-function list in which CUP was the alternative that would be chosen for DANGER on the test stage if *S* utilized the mediating chain. In all cases, the two alternatives for a given item had been of equal exposure during acquisition to reduce response bias. The four items in the unmixed list were identical to the chaining items in the mixed list.

Procedure.—The double-function list was presented as a paired-associate task using a study–test method. Three different orders of the study list were typed on a tape for presentation on a Stowe memory drum at a 2:2 sec. rate for study and test intervals, respectively, with a 2-sec. intertrial interval. The list was performed to a criterion of two errorless trials, not necessarily in a row. After a short interval, four orders of the mediation test list were presented on a tape, and an *S*-paced rate for presenting items was adopted. The list was continued until the same alternative for each item had been given on five consecutive trials. The instructions for this stage required *S* to say the stimulus word and then pick one of the two response words that he thought might be correct, remembering the first list that he learned. The *S* was required to give the number (1 or 2, representing the left–right position) of that word. The *S* was further told that he should change his answer in later presentations of the list only if he had a reason.

Subjects.—The 64 *S*s completing this experiment participated to fulfill a course requirement in an introductory psychology class at the University of Iowa. They were assigned randomly to two groups of equal size for mixed and unmixed presentation of the test list with the restriction that half were male and half were female in each group. These groups were further divided with one half receiving a form of the double-function list making a particular alternative the chaining alternative and the other half receiving the form making the other alternative the chaining alternative. Eighteen *S*s were discarded due to the inability to reach the acquisition criterion on the double-function list within 35 trials.

Experiment II

Materials.—The materials were similar to those used by Schulz et al. (1965) but are described briefly. The A-B list used in Stage 1 was composed of eight paralog-noun pairs. The paralogs were of low meaningfulness and the nouns of high meaningfulness. The B-C list used in Stage 2 was composed of eight noun-trigram pairs, the nouns coming from the Stage 1 list. The trigrams were low-meaningful CVCs. No paralog had as its highest free associate any of the subsequent nouns. Internoun association was minimal and formal similarity for CVCs was minimal. The A-C list used in Stage 3 was composed of the paralogs from Stage 1 and the trigrams from Stage 2.

Procedure.—The lists for Stages 1 and 2 were presented using a study–test method. During study trials the pairs were presented singly at a 2-sec. rate. During test trials a multiple-choice procedure was used, which consisted of presenting the stimulus term with four of the response terms numbered in order from one to four. The *S* was instructed to say the number corresponding to the appropriate alternative within a 2.5-sec. exposure duration. In assigning the distractors for the multiple-choice test trials, counterbalancing measures were used with regard to the frequency of occurrence and ordinal position among the alternatives. The intertrial interval was 2 sec., during which a yellow tape appeared in the aperture. Four random orders of the list were repeated until a criterion of two successive errorless test trials was reached.

Six random orders of the Stage 2 list were presented as a test-only task. A multiple-choice procedure was used which consisted of presenting the stimulus term with either two or three alternatives numbered in order. The instructions for Stage 3 were similar to the test-only instructions in Exp. I but the emphasis on choosing the same alternative unless *S* had a reason was dropped. The *S* was instructed in this stage to say the number corresponding to the alternative that "goes with" the paralog. The items were presented at a 3-sec. rate, but if *S* failed to respond, the drum was turned back to the item and *S* was required to choose one of the alternatives. The intertrial interval was 3 sec. For the two-alternative group the list was repeated until the same alternative had been chosen on five consecutive trials for each item. For the three-alternative group the list was repeated until the same alternative had been given on four consecutive trials.

Subjects.—The 48 *S*s completing this experiment had been assigned randomly to the two groups of equal size with the restriction that half were male and half were female in each group. The two-alternative group was subdivided into two groups, each receiving a different Stage 2 list causing one of the two alternatives to be the chaining alternative. Likewise, the three-alternative group was subdivided into three groups, each receiving a different Stage 2 list causing one of the three alternatives to be the chaining alternative. Five *S*s were discarded due to their inability to reach the acquisition criterion on the Stage 1 list within 35 trials.

RESULTS

Mediation Effects

The two groups in Exp. I did not differ significantly in the mean number of trials to criterion on the double-function list,

TABLE 1

RESULTS OF TESTS OF STATIONARITY AND INDEPENDENCE

Exp. I		Exp. II			
		Two alternative		Three alternative	
Chaining items	Nonchaining items	End in C	End in X	End in C	End in X
Chi-square test for stationarity					
$\chi^2(7) = 4.57$ $p > .50$	$\chi^2(7) = 1.33$ $p > .50$	$\chi^2(8) = 7.99$ $p > .40$	$\chi^2(6) = 7.19$ $p > .20$	$\chi^2(11) = 3.98$ $p > .50$	$\chi^2(8) = 5.42$ $p > .50$
Split-half difference test for stationarity					
$t(49) = 1.20$ $p > .20$	$t(30) = 1.19$ $p > .20$	$t(46) = .35$ $p > .50$	$t(20) = 1.81$ $.10 > p > .05$	$t(71) = 1.32$ $p > .10$	$t(35) = 1.43$ $p > .10$
Chi-square test for independence					
$\chi^2(1) = .77$ $p > .30$	$\chi^2(1) = .49$ $p > .40$	$\chi^2(1) = 2.04$ $p > .10$	$\chi^2(1) = 1.08$ $p > .20$	$\chi^2(1) = .04$ $p > .50$	$\chi^2(1) = .12$ $p > .30$

$F(1, 60) < 1$. Similarly, the two groups in Exp. II did not differ significantly on either Stage 1 or 2, $t(46) < 1$, in both cases. These results would indicate that any subsequent differences on the test stage may not be explained by appealing to differences in the paired-associate learning ability of the two groups in each experiment.

In order to establish mediated responding in a manner comparable to previous studies, only the first five trials of the test stage will be considered at the present. In the mixed-list group in Exp. I, the mean number of choices of the Chaining Alternative C was 3.09 per chaining item, whereas the mean number of choices of the corresponding Alternative X_1 was 2.33 per nonchaining item. Analysis of variance indicated that type of item (chaining vs. nonchaining) was a significant source of variance, $F(1, 28) = 6.61$, $p < .05$. In the unmixed-list group the mean number of choices of a chaining alternative was 3.38 per item. This value was significantly greater than the value expected by chance alone (2.50), $t(31) = 4.75$, $p < .001$, using a one-tailed test. The difference between the means for the chaining items in the mixed-list and unmixed-list groups was not significant, $t(62) = 1.16$.

In Exp. II the mean number of choices of the chaining alternative in the two-alternative list was 3.29 per item, which was significantly greater than the value expected by chance responding (2.50), $t(23) = 3.71$, $p < .01$. In the three-alternative list the mean per item was 2.73, which also was greater than the value expected by chance responding (1.67), $t(23) = 4.94$, $p < .001$. Thus, reliable mediated responding was found for each group in each experiment.

Tests of the All-or-None Assumption

The all-or-none assumption of the Markov model focuses on choice behavior on trials before entry into an absorbing state. The random model requires that during the random choice trials there be stationarity (the choice probabilities must be constant from one trial to the next) and independence of responses (the choice probability on Trial n must be statistically independent of the choice made on Trial $n - 1$). A chi-square test of stationarity, a split-half difference test of stationarity, and a chi-square test of independence were applied, where the model predicts nonsignificance for each test. These tests are explained in detail by Atkinson, Bower, and Crothers (1965, pp. 39–49). In the subject-item protocol the trials before the last switch in choice for a given item are assumed to represent random-choice trials. In Exp. I, items reached criterion early so that the number of observable random-choice trials was small. In Exp. II this was not the case, and a large number of random-choice trials was observed.

If the proportion of C alternatives chosen in the trials before criterion depends on the alternative in which the item ends, then response protocols for items ending in Alternative C must be analyzed separately from items ending in an X alternative. In Exp. I, no significant dependency on the alternative in which the item ended was indicated for the chaining and nonchaining items. Furthermore, the random-choice trials of the chaining items in the mixed and unmixed lists were combined to increase the power of the tests for stationarity and independence. In Exp. II, the proportion depended on the alternative in which the item ended in both the two-alternative and three-alternative groups so that the observations of the random-choice trials could not be combined.

The results of the tests of stationarity and independence are shown in Table 1. It can be seen from the p values that the model's assumptions of stationarity and independence of random-choice trials are tenable. The plausibility of the all-or-none assumption is indicated by the fact that of the 18 separate tests, none reached significance at the .05 level. Hence, further tests of the model's predictions were conducted. These tests, unlike the tests for stationarity and independence, required estimation of the model's parameters.

TABLE 2

ESTIMATES OF PARAMETERS

Estimate	Exp. I			Exp. II	
	Unmixed-list chaining items	Mixed-list chaining items	Mixed-list nonchaining items	Two-alternative group	Three-alternative group
p	.718	.675	.465	.730	.697
c	.342	.351	.290	.221	.154
$g_{c/c}$	.662	.517	.614	.655	.495
$g_{c/x}$	.381	.575	.373	.520	.204

TABLE 3

TESTS OF PARAMETER INVARIANCE BETWEEN ITEMS THAT END IN ALTERNATIVE C OR ALTERNATIVE X

Cond.	c free	c invariant
g_c free	.00	3.67 (1)
g_c invariant	90.08 (1)	91.61 (2)

Note.—Each entry is $-2(\max \log L)$ subject to the appropriate constraints minus $-2(\max \log L)$ subject to no constraints. Parenthesized numbers are the degrees of freedom of each chi-square relative to the cell with no constraints.

Parameter Estimates

Estimates of p, c, and g_c were obtained in the following manner: Since parameter p is defined as the probability of an item ending in Alternative C, it may be estimated directly from the observed proportions of items ending in Alternative C. These estimates are presented in Table 2. For c and g_c the present model states that the same parameters apply whether the item ends in Alternative C or in Alternative X. A more general model would include different parameter values for these two types of items. This model may be thought of as having one two-state process for items ending in Alternative C and another two-state process for items ending in Alternative X, where values of c and g_c could be different for the two processes. No transitions between the two processes would be allowed.

The simplifying assumptions of the present model that values of c and values of g_c be invariant for the two processes constitute a set of null hypotheses that can be evaluated by likelihood ratio tests. The tests of invariance of one parameter are not independent at different constraints of the other parameter. However, if all the possible tests are conducted to form a matrix of results, each null hypothesis (assumption of invariance for a particular parameter) can be evaluated in combination with all the possible hypotheses about the remaining parameters. This method of assessing parameter invariance is reported in greater detail by Greeno and Scandura (1966). Table 3 presents this matrix for all the data. In the first column, estimates of c are allowed to differ for items ending in Alternative C or Alternative X, and in the second column, c is assumed to be the same for both types of items. Analogously, in the first row g_c may differ for the two types of items and in the second row g_c must be the same. Each entry in the table is distributed asymptotically as χ^2 with the degrees of freedom as shown in parentheses determined by the number of constraints placed on the cell relative to the hypothesis of no constraints. As shown in Table 3, none of the test statistics for hypotheses that include invariance for c are significant relative to corresponding hypotheses that allow c to be free. However, separate estimates of g_c must be obtained depending on the alternative in which the item ends. This is reflected in the fact that the entries in the second row are statistically significant relative to those in the first row.

These tests indicate that two values of g_c must be estimated. Let $g_{c/c}$ be the parameter applying for items ending in Alternative C and $g_{c/x}$ be the parameter applying for items ending in Alternative X. The parameter c is invariant over items ending in Alternative C or Alternative X. This leads to a model which represents a compromise between the general model and the initially proposed model. Though the same transition rates hold for items ending in Alternative C and items ending in Alternative X, the response probabilities are different in the random-choice state. This may be represented by two different random-choice states, R_c for items ending in Alternative C and R_x for items ending in Alternative X, and is described by the following transition matrix and probability vectors:

$$\begin{array}{cc} & \text{State on Trial } n+1 \\ \begin{array}{c} \\ \text{State on} \\ \text{Trial } n \end{array} & \begin{array}{c|cccc} & C & R_c & R_x & X \\ \hline C & 1 & 0 & 0 & 0 \\ R_c & c & 1-c & 0 & 0 \\ R_x & 0 & 0 & 1-c & c \\ X & 0 & 0 & 0 & 1 \end{array} \end{array} \quad \begin{array}{c} \text{Start Vector} \\ \begin{bmatrix} pc \\ p(1-c) \\ (1-p)(1-c) \\ (1-p)c \end{bmatrix} \end{array} \quad \begin{array}{c} P(C) \\ \begin{bmatrix} 1 \\ g_{c/c} \\ g_{c/x} \\ 0 \end{bmatrix} \end{array}$$

The maximum likelihood estimates for the parameters c, $g_{c/c}$, and $g_{c/x}$ are presented in Table 2.

Effects of independent variables on the model's parameter.—The same type of analysis may be extended to testing parameter invariance over experimental conditions. Such an analysis will reflect the effects of independent variables on the model's parameters. Table 4 presents the matrices for the three experimental comparisons. The matrix of results for chaining and nonchaining items in the mixed-list condition in Exp. I may be interpreted as meaning that the parameters c, $g_{c/c}$, and $g_{c/x}$ do not vary significantly between the two conditions. The entries for hypotheses of invariance do not differ significantly from the corresponding entries for hypotheses that allow these parameters to be different. However, if the observed values of p are compared, a significant difference is found, $\chi^2(1) = 11.38$, $p < .001$. Thus, in terms of the model, the chaining and nonchaining items appear to be alike in all respects except that the proportion of chaining

TABLE 4

TESTS OF PARAMETER INVARIANCE BETWEEN EXPERIMENTAL CONDITIONS

	Chaining versus nonchaining		Mixed versus unmixed list		Two versus three alternative	
	c free	c invariant	c free	c invariant	c free	c invariant
$g_{c/c}$ and $g_{c/x}$ free	.00	.81 (1)	.00	.06 (1)	.00	10.97 (1)
$g_{c/c}$ and $g_{c/x}$ invariant	.58 (1)	1.46 (2)	6.58 (1)	7.14 (2)	59.54 (1)	74.67 (2)

Note.—Each entry in a particular matrix is $-2(\max \log L)$ subject to the appropriate constraints minus $-2(\max \log L)$ subject to no constraints. Parenthesized numbers are the degrees of freedom of each chi-square relative to the cell with no constraints.

items that end in Alternative C is greater than the proportion of nonchaining items than end in the corresponding Alternative X.

The matrix of results for chaining items in mixed- and unmixed-list conditions in Exp. I indicates that the parameter c can be assumed invariant between the two conditions but $g_{c/c}$ and $g_{c/x}$ differ. In the unmixed-list condition there appears to be a greater tendency in State R to choose the alternative in which the item ends. Finally, the values of p for the two conditions did not differ significantly.

From the matrix of results for constraints on the parameters over the two-alternative and three-alternative conditions in Exp. II, it appears that c, $g_{c/c}$, and $g_{c/x}$ all vary across conditions. The differences between the estimates of $g_{c/c}$ and $g_{c/x}$ for the two conditions certainly are expected since the number of alternatives varied. The values of c in Table 2 indicate that the probability of leaving the random-choice state on a given trial was less for the three-alternative condition than for the two-alternative condition. A standard correction for chance may be applied to the observed values of p in order to estimate the proportion of items on which S used the experimentally induced mediator. This estimate is .464 for the two-alternative condition and .546 for the three-alternative condition; however, they do not differ significantly.

DISCUSSION

The two experiments represent independent tests of the all-or-none model of verbal mediated responding. In both cases, the all-or-none assumption of the model appeared tenable since the predictions of stationarity and independence of responding in the random-choice trials were not rejected. However, maximum likelihood tests indicated that a slightly more general model than that initially proposed was needed. This model included separate estimates of g_c dependent on whether the item ended in Alternative C or Alternative X. Thus, it would appear that performance in the random-choice state is related to the probability of ending in Alternative C. In addition to the goodness-of-fit tests reported, theoretical and observed values of various statistics like the mean and variance of the trial number of the last switch in choice and the total number of choices of a particular alternative were compared.[3] Though minor discrepancies could be noted, they were not great enough to warrant any changes in the model's assumptions.

The assumption of an all-or-none process occurring in the choice of mediators is an important one. Though the mediators in the pool of possible associates may have attained their status due to their strength of association acquired during previous stages, the probability that a particular mediating chain of associations comes to be chosen consistently over trials appears to depend on an all-or-none process. It may be possible to relate this mathematical process to verbal hypotheses about the subprocesses of mediation. For example, the assumption of an all-or-none transition from State R to States C or X may be used to identify the concept of the "discovery" of a mediator. Schulz and Lovelace (1964) proposed the concepts of discovery and utilization of verbal mediators as being central to the understanding of mediational processes. Discovery was thought to be a function only of the study interval during paired-associate learning of A-C. However, this relationship could not be maintained due to the findings of Weaver et al. (1968) that study trials were not a necessary requirement for mediation to occur. Nevertheless, they maintain that the concept of discovery may still be useful. If an all-or-none process is implied by "discovery," then the concept may be anchored to a theoretical process that may be tested empirically. The present study represents the first attempt to test directly the notion of an all-or-none choice of a verbal mediator.

One advantage of a formal model is that the effects of independent variable manipulations often can be localized in one or two of the model's parameters. The results of the present analyses (as summarized in Table 4) would indicate that the effect of the establishment of an experimentally induced mediator (chaining vs. nonchaining conditions) is localized in the parameter p, the proportion of items ending in Alternative C. This implies that the magnitude of the mediation effect is best measured by p when c or g_c differ between conditions. When c and g_c are invariant, then the number of chaining alternatives chosen in a fixed number of trials would be a linear transformation of p and essentially would convey the same information. The parameter c was found to be greater for the two-alternative list than for the three-alternative list. It may be that it is the increase in information transmission with three alternatives over two alternatives that reduces the probability of giving the same alternative consistently. In paired-associate learning an increase in information transmission increases the latency of responding (Judd & Glaser, 1969). It is possible that in this task an increase in the time necessary to process the material is responsible for the decrease in c. Confirmation of this hypothesis must await experimentation in which the latencies are obtained. Finally, the parameter g_c appears to be affected primarily by the number of alternatives per item. That g_c differs for items that end in Alternative C or Alternative X might indicate an interdependence between the probability of choosing Alternative C in the random-choice state and the probability that S will come to choose Alternative C consistently.

---

[3] Copies of the table of theoretical and observed values and the equations used to calculate the predicted statistics are available from the first author.

There is some question as to whether the entire mediation effect is observed on the first trial or whether performance could increase over trials due to mediation. Greeno and Scandura (1966) used a paired-associate transfer task and found that all facilitation due to verbally mediated concepts was evidenced on the first trial or not at all. This was not the case in the present experiments. Increased mediated responding was observed on later trials. However, the present task differs in that no feedback is given to S as to whether he is correct or not. It is possible that in the paired-associate learning task the increase in correct responses due to reinforcement overshadows the facilitation due to transfer past the first trial. Weaver et al. (1968) argue that, in the paired-associate learning task, performance is confounded by study trials and the formation of nonmediated associations. In the test-only task, conditions are optimal for the investigation of mediated performance. It is interesting to note that in the present model all mediated responding would be evidenced on the first trial if either c were near one or g_c were equal to p.

Though the present results tend to support the model, it is not proposed that this type of Markov model be the final description of the process underlying mediated responding. Rather it is meant to be an initial attempt to test formally some basic assumptions about the mediational process underlying behavior in certain choice tasks. The present study demonstrates that a mathematical model may be usefully applied to mediation experiments. It also indicates that a mathematical process may be used to give formal meaning to a previously unanchored term like "discovery," and that the role of independent variables may be localized in terms of their effects on the parameters of the model.

REFERENCES

ATKINSON, R. C., BOWER, G. H., & CROTHERS, E. J. *An introduction to mathematical learning theory.* New York: Wiley, 1965.

GREENO, J. G., & SCANDURA, J. M. All-or-none transfer based on verbally mediated concepts. *Journal of Mathematical Psychology*, 1966, **3**, 388–411.

JAMES, C. T., & HAKES, D. T. Mediated transfer in a four-stage, stimulus-equivalence paradigm. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 89–93.

JUDD, W. A., & GLASER, R. Response latency as a function of training method, information level, acquisition, and overlearning. *Journal of Educational Psychology*, 1969, **60** (4, Pt. 2).

NORMAN, K. L., LEVIN, I. P., & WILLIAMS, J. R. Mediated responding on a multiple-choice test-only list following the acquisition of a double-function list. *Psychonomic Science*, 1970, **21**, 77–79.

PETERSON, M. J., & BLATTNER, K. C. Development of a verbal mediator. *Journal of Experimental Psychology*, 1963, **66**, 72–77.

SCHULZ, R. W., & LOVELACE, E. A. Mediation in verbal paired-associate learning: The role of temporal factors. *Psychonomic Science*, 1964, **1**, 95–97.

SCHULZ, R. W., WEAVER, G. E., & GINSBERG, S. Mediation with pseudomediation controlled: Chaining is not an artifact! *Psychonomic Science*, 1965, **2**, 169–170.

WEAVER, G. E., HOPKINS, R. H., & SCHULZ, R. W. The A-B, B-C, A-C mediation paradigm: A-C performance in the absence of study trials. *Journal of Experimental Psychology*, 1968, **77**, 670–675.

(Received November 20, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 255–262

HOMONYMS AND SYNONYMS AS RETRIEVAL CUES[1]

LEAH L. LIGHT[2]

Pitzer College

The two experiments reported here investigated the properties of effective retrieval cues. All *S*s studied a list of 45 to-be-remembered (TBR) homonyms. In Exp. I homonyms were studied alone or embedded in sentences; in Exp. II homonyms were studied as isolated units. Memory was tested by means of (*a*) free recall, (*b*) a two-alternative forced-choice recognition test, (*c*) a two-alternative forced-choice quasi-recognition test in which *S*s circled homonyms of TBR words, and (*d*) a two-alternative forced-choice quasi-recognition test in which *S*s circled synonyms of TBR words. Memory test scores were ordered (from poorest to best) from free recall, to synonym cues, to homonym cues, to recognition. Implications of these results for the encoding specificity hypothesis are discussed.

In cued recall experiments, *S*s are typically presented with a list of to-be-remembered (TBR) words to study. At some later time they are asked to recall the list and are given prompts or retrieval cues for the TBR words. Retrieval cues are said to be effective to the extent to which they result in recall scores higher than those obtained by unaided or free recall. The present article is concerned with the effectiveness of homonyms and synonyms as retrieval cues.

Thomson and Tulving (1970) have recently differentiated between two hypotheses regarding the nature of effective retrieval cues: the encoding specificity hypothesis and the associative continuity hypothesis. In their words,

> The encoding specificity hypothesis . . . clearly implies that no cue, however strongly associated with the TBR item or otherwise related to it, can be effective unless the TBR item is specificially encoded with respect to that cue at the time of its storage. . . . According to the associative continuity hypothesis, if a strong preexperimental association exists between Verbal Units A and B, then A can serve as an effective retrieval cue for B, and vice versa, simply by virtue of the existence of such an association and regardless of the specific nature of encoding events occurring at the time of the storage of the TBR unit [p. 255].

As is often the case, neither hypothesis alone seems to be capable of accounting for all the experimental findings. On the one hand, several studies have found that cues presented for the first time at recall are either ineffective in boosting recall scores or else have a positively detrimental effect on recall (e.g., Dong & Kintsch, 1968; Earhard, 1969; Funkhouser, 1968; Thomson & Tulving, 1970; Tulving & Osler, 1968). On the other hand, cues which are associatively or semantically related to TBR words have proven effective when presented for the first time at recall (e.g., Bahrick, 1969, 1970; Bilodeau & Blick, 1965; Lloyd, 1964; Wood, 1967). Also, Kintsch and Hogan (1971) have found that cues presented for the first time at recall were effective after several trials in a multitrial free-recall experiment.

One difficulty in evaluating this research stems from the fact that it is impossible to know how *S* is encoding the TBR words when they are presented. Therefore, results which fail to support the encoding specificity hypothesis can sometimes be discounted by arguing that *S*s may indeed have encoded TBR words with respect to retrieval cues later offered by *E*. This problem is less serious for hypotheses about the nature of effective retrieval cues offered by Bahrick (1969, 1970), Bower (1970), and Kintsch and Hogan (1971). Although

---

[1] This research was supported by a grant from the Research and Development Committee of Pitzer College. Roland Dumas, Charlotte Patterson, and Sharon Zweig assisted at various stages of the two experiments. The author is particularly grateful to L. Munter of the Claremont Unified School District, A. Hughes, and the teachers and students of the Claremont High School for their cooperation in Exp. I. Experiment II would have been impossible without the generous cooperation of M. P. Friedman.

[2] Requests for reprints should be sent to Leah L. Light, Pitzer College, Claremont, California 91711.

there are some differences among them, these hypotheses have in common the notion that recall involves both search and recognition. Retrieval cues initiate search in long-term memory with recognition checks being performed on retrieved items; items passing the recognition test are produced as overt responses. Let us assume that words are stored in long-term memory as lists or collections of units which specify their meaning, sound, and other attributes. Retrieval cues permit search to be initiated on the basis of some one or more of these properties. The items specified by a given cue may be called the search set (after Shiffrin, 1970). On this view, any word which includes a TBR word in its search set may serve as a retrieval cue for the TBR word. The probability that a TBR word will be included in the search set of some retrieval cue will be a function of the extent to which the retrieval cue and the TBR word share common properties as well as of the particular search strategy that *S* uses at a given moment. (For instance, if examination of the entire search set is not carried out, a TBR word may not be retrieved even if it shares many properties with some retrieval cue.) Hence, it is not necessary that *S* specifically encode a TBR word in respect to some retrieval cue at the time of storage in order for that cue to be effective. The only requirement is that there either be some previously established relationship between the TBR word and a cue or that *S* establish such a relationship through some mediational process during storage.

The two experiments reported here investigated the relationship between search set specificity and recall accuracy. As Shiffrin (1970) has pointed out, recognition memory tests are tasks in which *S* is given the maximum amount of information about search set—all the properties of the word are specified and the amount of search required before performing a recognition check is minimal. In the case of free recall, almost no information about search set is given at the time of recall so that the amount of search is at a maximum. Between these two extremes lie situations in which *E* offers retrieval cues which vary in the probability with which they include TBR words in their search sets. Although they typically overlap very little in semantic properties, homonym pairs such as ISLE and AISLE or ALLOWED and ALOUD have identical or nearly identical acoustic properties and are often highly similar in orthography. Thus, either member of a homonym pair ought to be an effective retrieval cue for the other when search is directed along acoustical properties of the cue because the overlap in relevant properties is high and the size of the search set is very small so that exhaustive examination of the search set is likely. However, while many English words have several synonyms, none of these synonyms is very exact—i.e., the overlap in semantic properties of synonym pairs is less than perfect. Synonyms should therefore be less effective as retrieval cues than homonyms because the overlap in semantic properties of synonym pairs is less than perfect and the size of the search set is likely to be relatively large so that the probability of an exhaustive search at the time of recall is reduced. Thus, one would predict a continuum of accuracy in performance from free recall, to synonym retrieval cues, to homonym retrieval cues, to recognition.

EXPERIMENT I

Method

Design.—This experiment investigated recall performance accuracy as a function of the nature of information about TBR words offered to *Ss* at the time of memory testing. All *Ss* studied a list of 45 homonyms and were then tested for memory by means of recognition, homonym retrieval cues, synonym retrieval cues, or free recall. Because words in isolation are often subject to more than one semantic interpretation, an effort was made to increase the likelihood that TBR words and their synonym cues would be given the same semantic interpretation by having half of the *Ss* study homonyms embedded in sentences which biased the semantic cue meanings of the TBR words. Sentence embedding was also designed to deter *Ss* from noticing that the words they were studying were all homonyms. The remaining *Ss* studied the TBR homonyms as individual words. The four types of memory test and the two methods of TBR word presentation generate eight experimental conditions: free recall-words (FR-W), synonym cues-words

(S–W), homonym cues–words (H–W), recognition–words (R–W), free recall–sentences (FR–S), synonym cues–sentences (S–S), homonym cues–sentences (H–S), and recognition–sentences (R–S).

Materials.—Forty-five homonym pairs were selected from Whitford (1966) such that intralist acoustic and semantic similarity were judged to be low. One member of each pair was chosen to be the TBR word and the other served as its homonym cue. A sample of 23 college-aged *S*s provided synonyms for the TBR words and the most frequent response for each homonym served as the synonym cue. On the average, each synonym was given as a response .51 of the time. The TBR words, homonyms, and synonyms were matched as closely as possible for frequency of occurrence. Examples of TBR words, their homonym cues, and their synonym cues include (in that order) HARE, HAIR, RABBIT; FAIRY, FERRY, ELF; SWEET, SUITE, SUGARY; CHIC, SHEIK, STYLISH; ALLOWED, ALOUD, PERMITTED; MANNER, MANOR, WAY; FISSURE, FISHER, CRACK. For each TBR word, a 5- to 10-word sentence was written which biased the semantic cue meaning of that TBR word; each sentence was composed so that the synonym cue could have been substituted into the sentence frame for the TBR word without undue awkwardness. Examples of such sentences are: He can run as fast as a HARE; the FAIRY lived in an enchanted castle; they had SWEET cakes with their tea; she wore a CHIC dress today; smoking is not ALLOWED here; he always does things in his own MANNER; the earthquake opened a FISSURE in the ground. The TBR words and sentences were dittoed, one word or sentence (with the TBR word underlined) per page and arranged in booklets in random order. The order of TBR words for sentence and word booklets was the same.

For the free-recall test, *S*s received a sheet of lined paper. The recognition test was a two-alternative forced-choice test; distractor items were homonyms roughly matched with the TBR words in length and frequency of occurrence. In the usual cued recall experiment, *S* is simply given a list of retrieval cues at the time of recall and told that they may or may not help him to remember the TBR words. In this experiment, some *S*s were to receive homonyms or synonyms of the TBR words as retrieval cues and were to be informed of the relationship between the cue and the TBR word. To prevent *S* from simply generating a homonym or a synonym for each retrieval cue in these conditions, a quasi-recognition procedure was employed. Each homonym cue was paired with the distractor used with the corresponding TBR word on the recognition test. In the H–W and H–S conditions, *S*'s task was to circle the word in each pair that sounded just like one of the TBR words and to write down the TBR word as well. Similarly, for the S–W and S–S conditions, each synonym was paired with another word and *S*'s task was to circle the one word in each pair that meant the same as a TBR word and to write down the TBR word. Distractors were matched roughly on length and frequency with synonym retrieval cues. For the recognition, synonym cues, and homonym cues tests, items were randomized so that each third of the test included equal numbers of items from the beginning, middle, and end of the study list. The order of items was the same for all three conditions.

Subjects.—All *S*s were high school students, in Grades 9–12, and were tested in intact classrooms.

Procedure.—In each classroom, half of the *S*s studied words and the remaining half studied sentences. Memory tests were randomized in blocks of four and were handed to *S*s in order of testing. Each *S* first received a study booklet with instructions printed on the cover. These instructions requested *S*s to study the words or sentences (paying particular attention to the underlined words) and informed them that there would be a memory test after the list had been studied. The nature of the memory test was not specified at this time. Every 5 sec. *S* turned a page of his booklet in response to a signal from *E* and studied its contents.

Following list study, booklets were collected and the written memory test was administered. Standard instructions were used for the free-recall and recognition conditions. Instructions for the H–W and H–S conditions requested *S* to circle the one word in each pair which sounded just like a TBR word and to write down the TBR word, gave examples of homonyms, and encouraged *S*s to guess if uncertain. Instructions for the S–W and S–S conditions asked *S* to circle the one word in each pair which meant the same as a TBR word and to write down the TBR word, gave examples of synonyms, and encouraged *S* to guess if uncertain. Unlimited time was allowed for the memory test.

Results

Although the experimental design called for equal numbers of *S*s in each cell, exigencies of testing in classrooms and the failure of some *S*s to follow instructions resulted in unequal cell frequencies. The numbers of *S*s lost due to failure to follow instructions were two from H–S, four from S–S, four from H–W, and two from S–W. All data were analyzed using unweighted means analyses of variance (Winer, 1962).

The data were analyzed in two ways. Scores for each *S* in the quasi-recognition conditions were first expressed as a proportion (of 45) of correct circling responses. Scores for *S*s in the recognition and free-recall conditions were also expressed as proportions correct. Proportions for *S*s in the recognition and quasi-recognition conditions were not corrected for guessing. Means and standard deviations for proportion correct as well as number of *S*s in each

TABLE 1

EXPERIMENT I: PROPORTION CORRECT ON RECOGNITION TASK

Statistic	Type of memory test			
	Recognition	Homonym cues	Synonym cues	Free recall
	Words			
M	.918	.838	.729	.319
SD	.075	.118	.116	.134
n	20	14	17	20
	Sentences			
M	.916	.806	.679	.180
SD	.059	.093	.092	.085
n	17	16	14	19

condition are given in Table 1. Scores evidently increase with cue specificity. Analysis of variance of these proportions resulted in an F (3, 129) = 298.92, $p < .001$, for type of memory test. The main effect for study method was also significant, F (1, 129) = 10.50, $p < .01$. Since the interaction was significant at the .05 level, F (3, 129) = 2.98, simple effects were computed for both factors. Simple effects for both types of study method were highly significant, with Fs of 121.19 and 180.71 for word and sentence conditions, respectively, both ps $< .001$. However, within type of memory test, only the difference between words and sentences on free recall reached significance, $F = 16.48$, $p < .001$, suggesting that the locus of the interaction effect is here. Newman-Keuls tests were performed on cell means within each type of study method. For both word and sentence study, all differences between cell means were significant at the .05 level.

Second, recall performance in the homonym cue, synonym cue, and free-recall conditions was compared to recognition performance. A strict scoring criterion was adopted: Credit was given for a word only if it was spelled correctly and was written next to its homonym or synonym in the quasi-recognition conditions. A laxer criterion would not have affected the outcome of the analyses. For this analysis, recall scores for the quasi-recognition conditions were conditionalized on correct circling—that is, instead of computing proportion correct out of 45 possible, proportion correct recalls *given correct circling* were computed—and were compared to free-recall and recognition scores. This step was taken because *S*s in the homonym cue conditions circled more words correctly than did *S*s in the synonym cue conditions and thus had a greater opportunity to use appropriate retrieval cues than did *S*s in the synonym cue conditions. Conditionalizing on correct circling ought to reduce this bias in analyzing the recall scores. Means of the conditional proportion correct recall scores and standard deviations are given in Table 2. Analysis of variance once again produced significant effects for method of study, F (1, 129) = 5.39, $p < .05$, and for type of memory test, F (3, 129) = 239.98, $p < .001$. The F (3, 129) for the interaction was 5.34 which is significant at the .01 level. Simple effects for word and sentence study were computed, giving Fs of 89.71 and 155.60, respectively, both ps $<$ .001. Again, study method failed to have an effect on recognition accuracy or on cue effectiveness for homonym cues, but affected synonym cue scores, F (1, 129) = 7.66, $p < .01$, and free-recall scores, F (1, 129) = 11.40, $p < .001$. Newman-Keuls tests performed on cell means within study conditions again revealed all means within each level to be significantly dif-

TABLE 2

EXPERIMENT I: CONDITIONAL PROPORTION RECALLED

Statistic	Type of memory test			
	Recognition	Homonym cues	Synonym cues	Free recall
	Words			
M	.918	.814	.510	.319
SD	.075	.150	.174	.134
	Sentences			
M	.916	.877	.396	.180
SD	.059	.069	.164	.085

ferent from each other at the .05 level with the single exception of the difference between H–S and R–S which was not significant.

The above analyses were also carried out using proportions to which the arc sine transformation had been applied. Analyses of the transformed data yielded virtually the same results as those of the original data and need not be reported.

Discussion

Regardless of whether one considers circling accuracy or recall performance in the quasi-recognition tasks, memory performance improves with retrieval cue specificity. As predicted, accuracy was best for recognition and poorest for recall with homonym and synonym cues producing intermediate performance. Groups cued with homonyms produced higher scores than did the free-recall groups. Similarly, groups cued with synonyms performed better than the free-recall groups; this finding is in line with Kintsch and Hogan's (1971) results. Nonetheless, homonym retrieval cues were far more effective than synonym retrieval cues. This result is in accord with findings by Shulman (1970) for short-term memory and with Klein's (1970) finding that false alarms on a recognition memory test following either long or short delays are much more likely to be homonyms of TBR words than semantically related words.

However, conclusions to be drawn from other aspects of the experiment are more ambiguous. It was predicted that the S–S condition would produce better recall than the S–W condition because *S*'s attention would be focused by the sentence context on semantic properties of the TBR words which were shared with the synonym cue. Although there was no difference between these two groups on the circling task, Group S–W performed significantly better on the recall task than did Group S–S. Similarly Group FR–W performed better than did Group FR–S. Why these two test conditions should have been sensitive to differences in study methods, while recognition and homonym retrieval cue conditions were unaffected by differences in study method, cannot be explained at this time.

EXPERIMENT II

Two possible objections to the inference that homonym cues are more effective than synonym cues result from methodological shortcomings of Exp. I. First, there was no way of evaluating whether the distractors used in the various conditions of Exp. I were equivalent in difficulty. If, for some reason, the distractors differed in attractiveness so that correct alternatives were more likely to be selected for circling in the recognition and/or homonym cue conditions than in the synonym cue condition, the results of Exp. I could be attributable to this artifact. Exp. II included three control conditions to assess the relative attractiveness of the distractors used in the various conditions. Second, *S*s in the homonym conditions in Exp. I might have become aware during list presentation that the TBR words were all homonyms. If they did become aware of the homonymous nature of the TBR words and encoded them specifically with regard to their homonym counterparts, the results of Exp. I could easily be attributed to encoding specificity. In Exp. II, all *S*s were asked at the end of the experiment whether they had noticed that the TBR words were homonyms so that performance of "unaware" and "aware" *S*s could be compared.

Method

Design and procedure.—Three experimental groups were involved in the main portion of this study. All *S*s studied the TBR words used in Exp. I. One group received a two-alternative forced-choice recognition test (Group R); one group received a synonym quasi-recognition test (Group S); and one group received a quasi-recognition homonym test (Group H). The *S*s were tested either individually or in small groups of two to four. In all other respects, these three groups were treated identically to the corresponding groups of Exp. I.

Three other control groups were later tested to determine whether or not guessing probabilities for the three types of tests were identical. The *S*s in the control experiment did not study the TBR list but simply took the memory test. They were instructed that other *S*s had studied a list of words. For the recognition control conditions, *S*s were told to select the one word in each pair that they believed to have been studied by the other *S*s. For the homonym and synonym control conditions, *S*s were asked to circle the homonym or the synonym of the word they believed the other *S*s to have studied. These *S*s were also tested individually or in groups of up to four *S*s.

Subjects.—The *S*s in the main experiment were college undergraduates whose participation in the

TABLE 3

EXPERIMENT II: PROPORTION CORRECT

Statistic	Type of memory test		
	Recognition	Homonym cues	Synonym cues
M	.951	.881	.671
SD	.066	.068	.070

experiment fulfilled a course requirement. Fifteen *S*s served in each experimental condition. Data from 1 additional *S* were discarded from the homonym condition for failure to follow instructions. The *S*s in the control experiment were primarily paid volunteers. Nine *S*s were tested in each control condition, but the data of 1 *S* in the synonym control condition who failed to follow instructions were discarded.

Results

As in Exp. I, the basic data to be analyzed were proportion correct recognition and quasi-recognition scores with no guessing correction applied. (See Table 3.) A one-way analysis of variance indicated significant differences among the three main experimental conditions, $F\ (2,\ 42) = 68.65$, $p < .001$. The proportion of variance attributable to experimental treatments was estimated by means of ω^2 to be .750. Performance was best for the recognition condition and worst for the synonym condition with the homonym condition falling inbetween; all pairwise comparisons of cell means were found to be significant at the .05 level by means of a Newman-Keuls test.

Means of the proportion correct for the three control groups were .551 for the recognition control, .544 for the homonym control, and .553 for the synonym control. Given the virtual identity of these means, it seems safe to conclude that the results of both Exp. I and Exp. II are free of artifacts having to do with differential "attractiveness" of distractors used in constructing memory tests in the recognition and quasi-recognition conditions.

The mean of the mean proportions correct for the three control groups was used as an estimate of base-line or "chance" performance for the three main experimental conditions. Proportion correct for each group was compared to chance, resulting in *t*s of 6.70, 18.86, and 23.68 for the synonym cue, homonym cue, and recognition conditions, respectively. All *t*s(14) were significant at the .001 level. Although Exp. II did not include a free-recall condition, these results suggest that both synonyms and homonyms may serve as effective retrieval cues.

GENERAL DISCUSSION

In the two experiments reported here, performance was best for recognition, intermediate for homonym and synonym cues, and poorest for free recall. These results do not support the specificity hypothesis. Most damaging is the finding that both homonyms and synonyms were quite effective as retrieval cues although these cues were never presented at the time of TBR word storage and first appeared at the time of output. Proponents of the encoding specificity hypothesis might claim that homonyms are such effective retrieval cues because their sounds are encoded at the same time as TBR words since they are acoustically and visually almost identical to the TBR words. To quote Thomson and Tulving (1970), a homonym cue might be said to correspond "to a part of the total encoding pattern representing the perceptual cognitive registration of the occurrence [p. 261]" of the TBR word during storage. The difficulty here is in determining the precise interpretation of the expression "a part of the total encoding pattern." If the encoding specificity hypothesis is interpreted to mean that *S*s must somehow explicitly consider a TBR word and a cue word together during storage for that cue to be effective later, homonyms should have been effective cues only for *S*s who noticed that the TBR words were homonyms during list storage. In Exp. II, *S*s were asked to indicate whether or not they had noticed that all or some of the TBR words were homonyms at the time they studied those words. In the homonym condition, 6 *S*s were classed as aware and 9 were classed as unaware on the basis of their answers to this question. Although aware *S*s scored a mean of .918 correct as compared to .857 correct for unaware *S*s, $t\ (13) = 1.85$, $p < .05$, one-tailed test, the unaware *S*s' mean score is still far higher than that of the synonym condition. Furthermore, a subsequent sample of 12 *S*s studied the homonyms embedded in the sentences used in Exp. I and were queried as to whether or not they had noticed the homony mous nature of the TBR words. None of these 12 *S*s could be classified as aware, suggesting that the results obtained for sentence study in Exp. I are not likely to be due to specific encoding of the homonymous counterparts of the TBR words.

Therefore, the encoding specificity hypothesis, as stated by Thomson and Tulving (1970), stands in need of revision. Consideration of the finding that homonyms and synonyms are effective retrieval cues may afford suggestions as to directions which such revision might take.

Homonyms and synonyms differ in at least two ways. First, homonyms are typically very similar in terms of acoustic and orthographic properties whereas synonyms are probably somewhat less similar in terms of semantic properties and are rarely similar in orthography. Such similarity may affect recall during either the retrieval or the recognition stages. During retrieval, *S* may limit search to items which are highly similar in properties to the retrieval cues available to him; hence, TBR words which are less similar to retrieval cues may not be retrieved at all (see Bahrick, 1969, 1970). On this argument, homonyms are more likely to be retrieved than synonyms. Alternatively, similarity may affect the recognition process directly. Retrieved words may be more likely to pass tests for the presence of recency or contextual tags if they share many properties with their retrieval cues. Such a notion would also account for the fact that intrusion errors in free recall of categorized lists tend to be members of those categories represented on the list rather than unrelated words (Bousfield, 1953), as well as for the fact that false alarms in recognition tend to be semantically or acoustically similar to TBR words (Anisfeld, 1969).

Second, homonym sets are no doubt smaller in size than synonym sets. Most homonym sets do not exceed three or four in number whereas synonym sets are probably considerably larger though their exact size would be difficult to estimate. Search set size may affect the probability of retrieval of TBR words in experiments such as those reported here. When confronted with homonym cues in the quasi-recognition conditions, *S*s may have been able to generate the entire search sets of both alternatives. Under these circumstances, recall would be dependent only on *S*'s ability to perform adequate recognition checks. On the other hand, when confronted with synonym cues in the quasi-recognition condition of the present experiments, *S*s face two potential problems: (*a*) If the search sets are large *S* may terminate search before retrieving a TBR word and be forced to guess, and (*b*) since the number of retrieved items would tend to be larger than in the homonym condition, *S* may experience difficulties in discriminating among alternatives even if the TBR word has been retrieved (Davis, Sutherland, & Judd, 1961). Still, synonym cues restrict memory search more than do free-recall instructions, and performance is better when synonym cues are given. Of course, the superiority of recognition as a measure of retention can be attributed either to the fact that search is limited to only one item or to the fact that the TBR word and its retrieval cue are identical.

Unfortunately, the present experiments do not permit a clear separation of the effects of similarity and search set size on the effectiveness of retrieval cues. Such a test would require that homonyms and synonyms be equated for degree of similarity and this condition was not met here. It is, however, possible to hold search set size constant by considering only the homonym pairs. Juola, Fischler, Wood, and Atkinson (1971) have suggested that homonym pairs which contain the same number of letters and have not more than two different letters be considered visually similar while those not meeting this requirement be considered visually dissimilar. The homonym pairs used in the present studies were classified as similar or dissimilar on this criterion. Each *S* was assigned a score representing the difference between the proportion of similar homonyms which were not recognized and the proportion of dissimilar homonyms which were not recognized. In Exp. I, *S*s were more accurate on visually similar homonyms, $t\ (29) = 2.31$, $p < .05$, for Groups H–S and H–W combined; in Exp. II, the difference was of marginal significance, $t\ (14) = 1.86$, $p < .10$. Thus, there is some evidence, albeit of a limited nature, that when search set size is held constant the efficacy of retrieval cues is a function of their visual similarity to TBR words. Future experimental work on the encoding specificity hypothesis should investigate the effects of search set size and degree of similarity of TBR words to retrieval cues. In this way, boundary conditions for the operation of the encoding specificity hypothesis may be established. Encoding specificity may be unnecessary when retrieval cues specify very small search sets or when the degree of similarity between cues and TBR words is high.

but when one or both of these conditions is not met, specific encoding may be required to ensure retrieval cue efficacy.

REFERENCES

ANISFELD, M. False recognition produced by semantic and phonetic relations under two presentation rates. *Psychonomic Science*, 1969, **17**, 366–367.

BAHRICK, H. P. Measurement of memory by prompted recall. *Journal of Experimental Psychology*, 1969, **79**, 213–219.

BAHRICK, H. P. Two-phase model for prompted recall. *Psychological Review*, 1970, **77**, 215–222.

BILODEAU, E. A., & BLICK, K. A. Courses of misrecall over long-term retention intervals as related to strength of pre-experimental habits of word association. *Psychological Reports*, 1965, **16**, 1173–1192. (Monograph Supplement)

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 1953, **49**, 229–240.

BOWER, G. H. Organizational factors in memory. *Cognitive Psychology*, 1970, **1**, 18–46.

DAVIS, R., SUTHERLAND, N. S., & JUDD, B. R. Information content in recognition and recall. *Journal of Experimental Psychology*, 1961, **61**, 422–429.

DONG, T., & KINTSCH, W. Subjective retrieval cues in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 813–816.

EARHARD, M. Storage and retrieval of words encoded in memory. *Journal of Experimental Psychology*, 1969, **80**, 412–418.

FUNKHOUSER, G. R. Effects of differential encoding on recall. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 1016–1023.

JUOLA, J. F., FISCHLER, I., WOOD, C. T., & ATKINSON, R. C. Recognition time for information stored in long-term memory. *Psychonomic Science*, 1971, **10**, 8–14.

KINTSCH, W., & HOGAN, R. M. Is specific encoding a prerequisite for the effectiveness of a recall cue? (Tech. Rep. CLIPR-4) Boulder: University of Colorado, 1971.

KLEIN, G. A. Temporal changes in acoustic and semantic confusion effects. *Journal of Experimental Psychology*, 1970, **86**, 236–240.

LLOYD, K. E. Short-term retention as a function of recall point coding. *Psychological Reports*, 1964, **14**, 752–754.

SHIFFRIN, R. M. Memory search. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

SHULMAN, H. G. Encoding and retention of semantic and phonemic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 499–508.

THOMSON, D. M., & TULVING, E. Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, 1970, **86**, 255–262.

TULVING, E., & OSLER, S. Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology*, 1968, **77**, 593–601.

WHITFORD, H. C. *A dictionary of American homophones and homographs*. New York: Teachers College Press, 1966.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

WOOD, G. Category names as cues for the recall of category instances. *Psychonomic Science*, 1967, **9**, 323–324.

(Received December 18, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 263–272

VESTIBULAR ADAPTATION IN MAN:

EFFECTS OF INCREASED ACCELERATION DURING DIFFERENT PHASES OF ADAPTATION[1]

TRACY LUSTER[2]

San Jose State College

The purpose of this study was to investigate the course of vestibular adaptation following a temporary increase in angular acceleration (*A*). Eight normal *O*s, who were rotated in a precision rotation device, gave periodic magnitude estimates of their perceived velocity throughout a prolonged 1°/sec² *A*, which was briefly increased to either 2° or 3°/sec² at four points in the adaptation cycle. The data, tested with analysis of variance, showed that the size of the magnitude estimates was influenced by the magnitude and position of the altered *A*. The duration of subjective response was affected only by the position of the altered *A*. Habituation, overshoot, and negative aftereffects were also observed. Response measures were shown to be interrelated. The results were discussed in reference to cupular displacement and vestibular adaptation.

The torsion pendulum model, based on cupular mechanics, successfully predicts subjective vestibular response to angular acceleration (*A*) when experimental conditions are similar to those of the natural environment (Guedry, 1965). Research has clearly established, however, that subjective response systematically diminishes below theoretical expectation and eventually disappears when *O* is exposed to prolonged, constant *A*s (Brown, 1966, 1968; Clark & Stewart, 1968, 1970; Parsons, 1970).

No reported studies have indicated how a change in *A* affects subjective response to a subsequent prolonged, constant *A*. This study was addressed to the problem of determining the effects that changing *A*s of different magnitudes and at different positions in the adaptation cycle have upon the perception of a constant *A*. Specifically, the hypotheses were: (*a*) There would be a systematic relationship between the position in the adaptation cycle during which a constant stimulus is altered and the subsequent adaptation to a constant stimulus, and (*b*) there would be a systematic relationship between the magnitude of the stimulus change and the overall adaptation rate to a subsequent constant *A*.

METHOD

Observers

Eight men between 21 and 31 yr. of age participated in this experiment. Their good health and normal vestibular and auditory functioning were established by self-affirmation and a physical examination. All had vestibular thresholds below .9°/sec², as determined by a prior study (Roberts, 1971). They had no previous experience with prolonged *A*s nor with the magnitude estimation procedure.

Apparatus

The Ames Man-Carrying Rotation Device (MCRD), a one-degree-of-freedom simulator capable of reaching velocities of 45 rpm in either direction, was used to present the *A*s (Clark & Stewart, 1968). The MCRD is a small, enclosed room with a comfortable chair equipped with a lap and shoulder harness and a headrest that holds *O*'s head stationary at the center of rotation. An analog computer controls the *A* magnitude of the MCRD within .01°/sec², and effects *A* changes of the MCRD within .1 sec. Distracting stimulation is reduced by the virtually vibration-free operation of the MCRD and by ventilating fans which partially mask external noise. An ink-writing recorder registers the velocity, acceleration, and computer output data.

---

[1] The author wishes to express his grateful appreciation to Brant Clark, John Stewart, and Edward Minium for their valuable guidance and assistance in the development and completion of this study. A special thanks is due Brant Clark, who made this study possible at Ames Research Center through NASA Grant NGL 05 046 002.

[2] Requests for reprints should be sent to Tracy Luster, who is now at the Department of Psychology, University of Nebraska, Lincoln, Nebraska 68508.

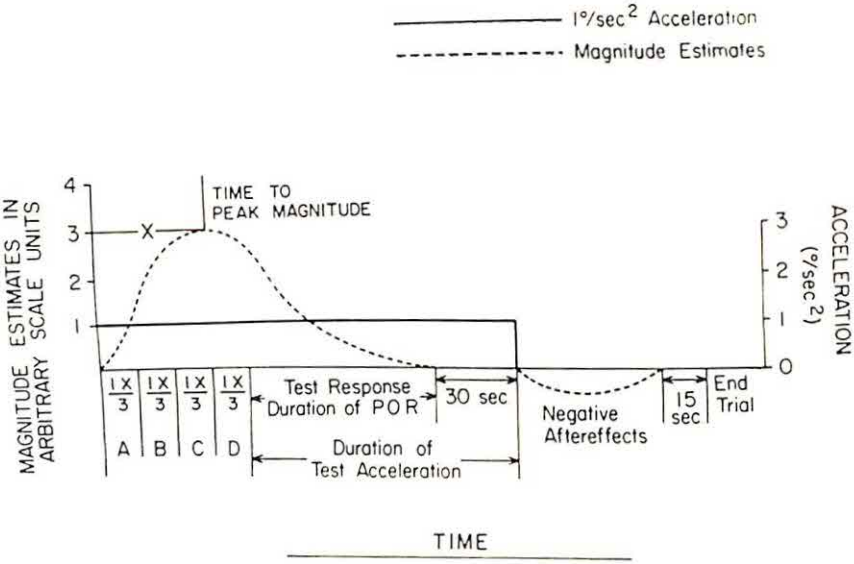


FIG. 1. Example of a base-line trial, stylized magnitude estimates, and the increments used to alter and present the independent variables in the experimental conditions.

Procedure

Observer's task.—The *O*'s task was to estimate the magnitude and direction of perceived rotation (POR) while in complete darkness. Throughout each trial, *O* verbally reported magnitude estimates at least once every 10 sec. and even more often while his magnitude estimates were changing. The estimates were recorded to the nearest second. The magnitude-estimate method, described in detail by Clark and Stewart (1968), involved a ratio scaling procedure in which *O* used his own numerical scale. He initially identified his chosen base number (modulus) with a minimum POR, and then reported proportionately greater numbers with corresponding increases of POR velocity.

Description of trials.—The *O* was accelerated to 10 rpm at the beginning of each session so that the rotational velocity attained during the first trial would not exceed the limits of the MCRD. A constant velocity was maintained for 2 min. before the first trial so that any aftereffects could dissipate. The *A* was then presented and *O* continually reported magnitude estimates throughout the trial. The *A* was terminated 30 sec. after *O* reported a null magnitude estimate followed by either another null magnitude estimate or a magnitude estimate in the opposite direction of the *A*. The magnitude estimates of any negative aftereffects were recorded, and the trial was terminated when *O* had given 15 sec. of consecutive null magnitude estimates. The control trial, during which there was no *A*, was terminated 2 min. after it began.

Sequence of trials.—Each of the two to four 30-min. trial sessions *O*s participated in each day consisted of six or seven trials with a 1-min. rest interval between each trial, and a 30-min. rest period between each session. There were three types of trial sessions: practice, base line, and experimental.

The first session consisted of six practice trials to familiarize *O* with the task, to allow him an opportunity to select a modulus for his magnitude estimates, and to allow him to choose his *POR* criteria for null magnitude estimates. He was instructed to use the same modulus and criteria throughout the remaining trials.

The second session consisted of one control and five base-line trials during which a prolonged, 1°/sec², constant *A* was presented. The *O*'s magnitude estimates of the base-line trials were used to determine the duration of the independent variable and the point in time at which it was to be introduced in the experimental trials. The magnitude estimates of the five base-line trials were used to determine the average time (*X*) needed for *O* to reach his maximum subjective velocity while exposed to the constant stimulus. If *O* gave two or more consecutive peak magnitude estimates, the midpoint between the first and last was designated as the point of peak subjective velocity for that trial. One-third of *X* was used to divide the early part of the adaptation cycle into four successive equal time segments for each *O*, beginning with the onset of *A* (Fig. 1). These periods, designated A, B, C, and D, cover the rise and initial decline of subjective velocity during a prolonged, 1°/sec², constant *A*.

The remaining sessions consisted of experimental trials during which a prolonged, 1°/sec² *A* was increased during Interval A, B, C, or D to an *A* of 2° or 3°/sec² (Fig. 2). There was also an additional experimental condition, identical to the base-line condition, in which the *A* was an uninterrupted constant 1°/sec², prolonged *A*. The nine experimental conditions and one control condition were

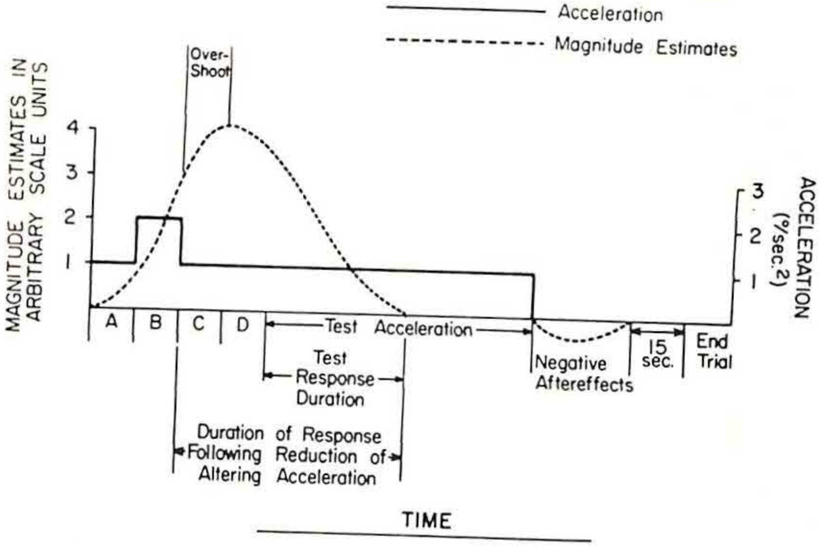


FIG. 2. Example of stylized magnitude estimates and the stimuli presented during an experimental trial in which a 1°/sec² *A* was increased to a 2°/sec² *A* during Interval B.

presented randomly in each of five blocks randomly ordered for each *O*.

RESULTS

The data were analyzed graphically and statistically. The Results section will include data for the following effects: the general profile of magnitude estimates plotted over time for the various conditions, the peak magnitude estimates, the differences between the peak magnitude estimates of the altered conditions and the corresponding magnitude estimates for the experimental constant *A* condition (simultaneous estimate differences), and response duration. Correlation matrices will also be presented to show the relationship between *O*'s unique length of altered interval and subsequent response measures and to show the interrelationship between response measures.

Group curves of the magnitude estimates of the positive effect.—A group curve representing the magnitude estimates was plotted over time for the positive effect of the base-line condition in a manner similar to that described by Clark and Stewart (1968). First, because *O* selected his own modulus, each *O*'s magnitude estimates for the five trials were divided by his modulus. The median magnitude estimate at each 5-sec. interval was determined from plots of the data for each *O*. The mean of the medians for the eight *O*s was computed and then plotted over time (Fig. 3). The *O*s' median time needed to reach the peak magnitude estimate ranged from 18 to 51 sec. The graph in Fig. 3 shows that the average time needed to reach the peak magnitude estimate during the base-line condition was approximately 25 sec. and that POR terminated 85 sec. after the beginning of the *A*.

The magnitude estimates for those experimental trials in which only a constant *A* was presented were plotted in the same manner as those of the base-line condition (Fig. 3). These data indicate that although the magnitude estimates were greater in the base-line condition than in the experimental, constant *A* condition, the overall response patterns were very similar. The curves also indicate that the time needed to reach the peak magnitude estimate was approximately 25 sec. for both conditions. That is an important observation since it validates the use of the duration of increasing magnitude estimates during the base-line trials as an appropriate interval for determining the duration and position of the independent variables in the various experimental, altered conditions.

The magnitude estimates for all the ex-

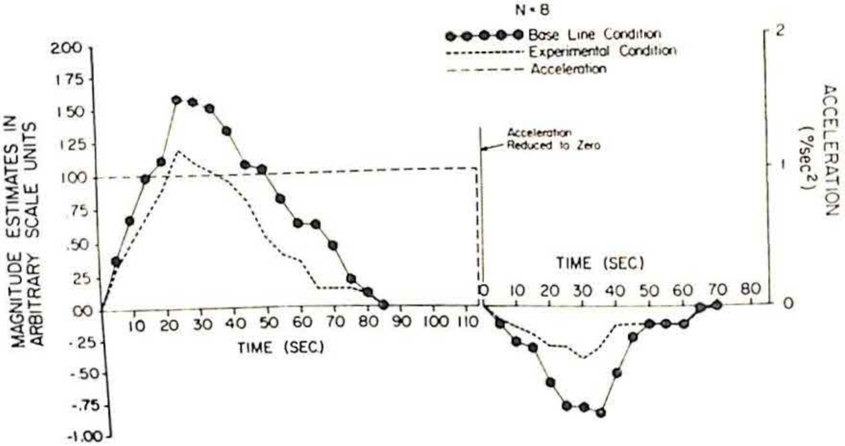


FIG. 3. Average magnitude estimates of rotational velocity during and following a 1°/sec² A for base-line and experimental, constant A conditions. (Estimates in the direction of the A are positive and those in the opposite direction are negative. Negative estimates describe negative aftereffect.)

perimental conditions were plotted in a manner similar to those shown in Fig. 3. The time increments, however, were standardized because the position and duration of the altered interval was unique for each *O*. The time increments for each *O*'s magnitude estimates were normalized by dividing *X* into 10 equal segments. The *O*'s adjusted magnitude estimates were then plotted at each normalized interval for each trial. The median magnitude estimates for each interval were then combined across *O*s, and the group mean magnitude estimate curves were plotted over normalized time increments for each condition (Fig. 4 and 5). The plots were terminated at 1.8*X* to avoid spurious end points because the depicted termination of the POR for the various conditions would have been a function of *X* rather than a representation of veridical response duration.

Peak magnitude estimates.—The peak magnitude estimates for each condition were combined across *O*s, and their means were then calculated (Table 1). The mean peak magnitude estimates are slightly different from those depicted in the curves in Fig. 4 and 5 because the peak magnitude

TABLE 1

MEAN PEAK MAGNITUDE ESTIMATES AND SIMULTANEOUS ESTIMATE DIFFERENCES, IN ARBITRARY SCALE UNITS, ACROSS EIGHT OBSERVERS FOR EIGHT ALTERED CONDITIONS

Response measure	Altered acceleration (°/sec²)	Altered interval			
		A	B	C	D
Peak magnitude estimate	2	1.68	2.19	2.22	1.91
	3	2.35	2.72	2.90	2.81
Simultaneous estimate difference	2	.78	.86	1.26	.72
	3	1.15	1.68	1.99	2.25

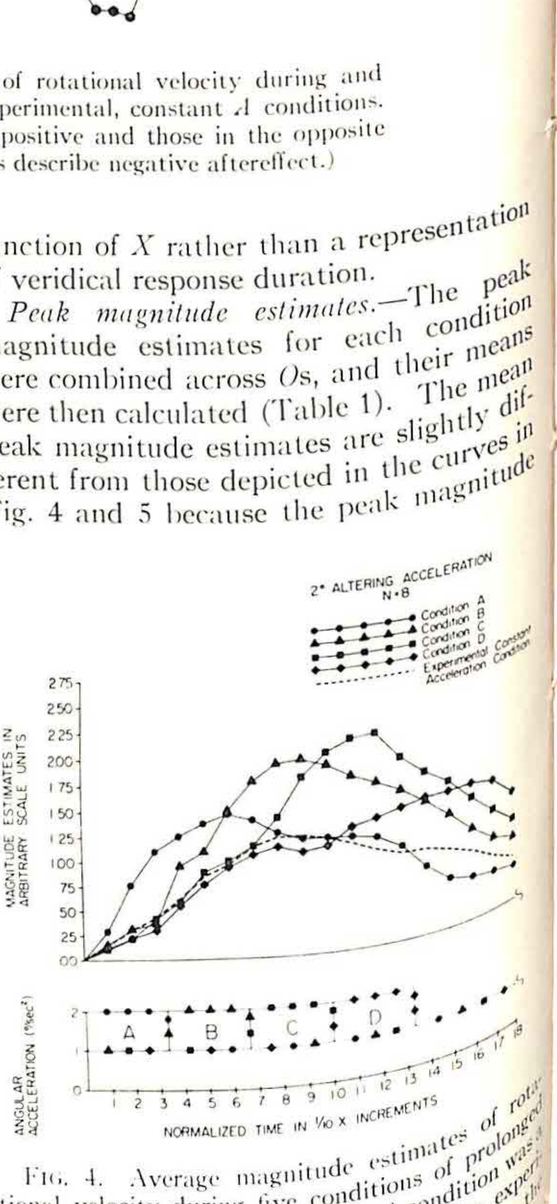


FIG. 4. Average magnitude estimates of rotational velocity during five conditions of prolonged A. (The experimental, constant A condition was a prolonged 1°/sec² A presented during the experimental sessions. The A presented during the altered interval was 2°/sec².)

estimates in the table represent the means of the mean magnitude estimates for each *O*, whereas those depicted in the curves are based upon the means of the median magnitude estimates for each *O*. Despite the different values, the magnitude estimates in Table 1 reflect the same pattern as those depicted in the group curves. The data in Fig. 4 and 5 and in Table 1 indicate that the manipulated variables had a systematic effect upon the size of the peak magnitude estimates. The effects that the variables had upon the size of the mean peak magnitude estimates were tested with a four-factor analysis of variance (Edwards, 1968). The four factors were (*a*) the magnitude of the increased *A*, (*b*) the interval during which the *A* was altered, (*c*) the trial block in which the condition occurred, and (*d*) the *O*. The results indicated that the mean peak magnitude estimates of the 3° altered conditions were significantly greater than those of the 2° altered conditions, $F\ (1,\ 7) = 38.51$, $p < .001$, and that the peak magnitude estimates became progressively smaller across subsequent trial blocks for all conditions, $F\ (4,\ 28) = 4.72$, $p < .01$. The results also showed that the point in time that the *A* was altered had a significant effect upon the size of the peak magnitude estimates, $F\ (3,\ 21) = 7.72$, $p < .01$. The variance attributed to *O*s could not be tested for significance because there was no appropriate error variance term (McNemar, 1962). There was no significant interaction between the three variables tested above, $p > .05$.

The data in Table 1 show that the peak magnitude estimates for Interval D in the 2° condition were substantially less than those for Interval C. The difference was significant when tested with a *t* test, $t\ (39) = 2.06$, $p < .05$.

Simultaneous estimate differences.—The difference between the peak magnitude estimate of each altered trial and the magnitude estimate reported at the corresponding time during the experimental, constant *A* condition was called the simultaneous estimate difference. The following steps were taken to determine the mean simultaneous estimate difference for each condition. The median magnitude estimate in the experimental, constant *A* condition reported at the same point in time as the peak magnitude estimate in each altered trial was first determined. The magnitude estimate in the constant *A* condition was then subtracted from the corresponding peak magnitude estimate of the altered trial. The median difference for the five trials of each condition was then calculated and combined across *O*s (Table 1). The effect that the independent variables had upon the simultaneous estimate difference was tested with a three-factor analysis of variance. The estimate differences were greater for the 3° conditions than for the 2° conditions, $F\ (1,\ 7) = 57.20$, $p < .001$, and the differences were generally greater as the *A* was altered later in the trial, $F\ (3,\ 21) = 4.66$, $p < .05$. Unlike the peak magnitude estimate measure, there was significant interaction between the magnitude and position of the increased *A*, $F\ (3,\ 21) = 5.57$, $p < .01$. Inspection of the

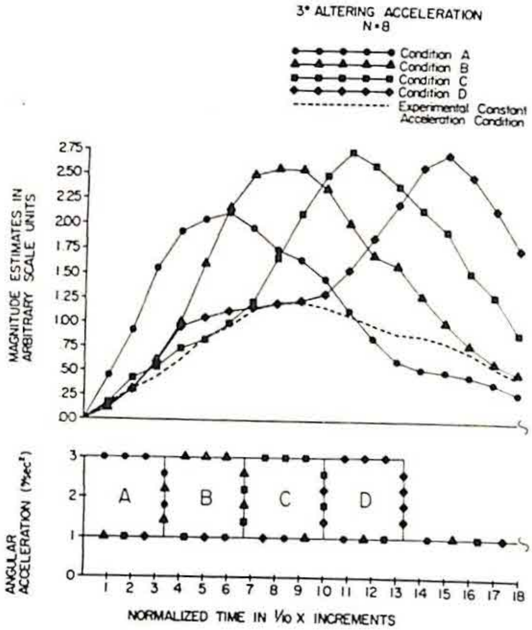


FIG. 5. Average magnitude estimates of rotational velocity during five conditions of prolonged *A*. (The experimental, constant *A* condition was a prolonged 1°/sec² *A* presented during the experimental sessions. The *A* presented during the altered interval was 3°/sec².)

TABLE 2

MEAN DURATION OF OVERSHOOT, TEST RESPONSE, AND POSTREDUCTION RESPONSE FOR EIGHT OBSERVERS

Response measure	Altered acceleration (°/sec²)	Altered interval			
		A	B	C	D
Overshoot effect	2	13.25	8.13	3.88	2.00
	3	9.63	4.75	5.50	4.63
Test response	2	10.15	19.45	22.58	29.55
	3	9.63	19.03	31.40	37.98
Postreduction response	2	43.13	42.13	34.13	33.00
	3	39.50	41.13	42.13	37.75

Note.—Mean duration is measured in seconds.

data in Table 1 shows that the significant interaction was apparently caused by the steeper increase of simultaneous estimate differences across the four altered intervals in the 3° condition, as compared to the 2° condition. Also, unlike the peak magnitude estimates, the estimate difference was greater for Interval D than for Interval C in the 3° condition.

Overshoot duration.—The increase in the size of magnitude estimates after the altered A was reduced to 1°/sec² has been called overshoot since the heavily damped torsion pendulum model does not predict an increase in perceived angular velocity after the A is reduced (Parsons, 1970). Substantial overshoot occurred in the various conditions of this experiment (Fig. 4 and 5). The average overshoot duration for each condition is shown in Table 2.

The effects that the magnitude of altered A, the altered interval, and O had upon overshoot duration were analyzed with a three-way analysis of variance. The data indicated that the duration of overshoot became progressively less as the A was altered later in the trial, $F\ (3, 21) = 21.31$, $p < .05$. The results also showed that overshoot duration was not significantly affected by the magnitude of the altered A and that there was no significant interaction between the magnitude and position of the altering A, $p > .05$. The variance attributed to O lacked an appropriate error variance term.

Duration of subjective response.—The duration of test response was arbitrarily selected as the duration of the POR following the interval $4X/3$ (Fig. 2). This measure was used in place of overall response duration to avoid experimental artifact caused by the variables manipulated during interval $4X/3$. The mean test response durations for the various conditions are shown in Table 2. A three-way analysis of variance was used to test for significant differences. The results indicated that the test response duration became progressively longer as A was altered later in the trial, $F\ (3, 21) = 15.88$, $p < .01$. Test response duration was not affected significantly by the magnitude of the increased A, nor was there a significant interaction between the size of the altered A and the position of the altered interval, $p > .05$.

The postreduction response duration was defined as the time between the reduction of the increased A and the first null magnitude estimate that O reported (Fig. 2). This duration measured the duration of subjective response following maximum cupular deflection for the altered conditions. The group means are listed in Table 2. The data were tested by analysis of variance. The results indicated that neither the magnitude nor the position of the altering A had a significant effect on postreduction response duration, $p > .05$. There was no significant interaction between these two variables, $p > .05$.

Negative aftereffects.—Magnitude estimates that were reported in the opposite direction to the preceding A and which began within 30 sec. after the termination of the A were recorded as negative aftereffects. Negative aftereffects were reported in 52% of the experimental trials. A four-way analysis of variance was used to test the effect the manipulated variables had upon the peak magnitude estimates and the duration of negative aftereffects. The peak magnitude estimates and response duration of the aftereffects were not significantly affected by the magnitude or position of the altered A nor by the trial block in which the condition occurred, $p > .05$.

Variable interrelationships.—The correlation matrix in Table 3 shows the correspondence between the unique length of each O's altered interval and the respon

TABLE 3

LINEAR CORRELATIONS BETWEEN OBSERVERS' LENGTH OF ALTERED INTERVAL AND FIVE RESPONSE MEASURES FOR ALL ALTERED TRIALS

Response measure	Altered acceleration (°/sec²)	Altered interval			
		A	B	C	D
Peak magnitude estimate	2	−.18	−.08	−.14	−.18
	3	−.04	−.13	−.21	−.13
Simultaneous estimate difference	2	−.02	.15	.24	.08
	3	.17	.08	.53***	.14
Overshoot duration	2	.40**	.11	.42**	−.01
	3	.09	.22	.35*	.09
Test response duration	2	−.32*	−.28	.02	.23
	3	−.42**	−.39*	−.08	.21
Postreduction response duration	2	.41**	.14	.23	.23
	3	.11	.09	.14	.21

* $p < .05$.
** $p < .01$.
*** $p < .001$.

measures for each of the eight altered conditions. Overall, the results indicate an insignificant or weak correspondence between the length of increased *A* and the various response measures. The duration of each *O*'s altered interval was not related to the size of the peak magnitude estimates nor, for the most part, to the simultaneous estimate differences, $p > .05$. The duration of altered *A* was significantly related to the simultaneous estimate difference in one condition, to overshoot duration in three conditions, to test response duration in three conditions, and to postreduction response duration in one condition.

The correlation matrix in Table 4 shows the interrelationship between response measures for combined altered conditions. Generally, the correlations (*r*s) in Table 4 indicate that the response measures reported in this experiment were not independent. There was a high correspondence between peak magnitude estimates and simultaneous estimate differences, and between test response duration and postreduction response duration. Overshoot duration was inversely related to the size of the peak magnitude estimates, but there was no significant relationship between overshoot duration and simultaneous estimate differences, $p > .05$.

The data in Table 4 include partial *r*s

TABLE 4

PARTIAL AND ZERO-ORDER LINEAR CORRELATIONS BETWEEN FIVE RESPONSE MEASURES FOR ALL ALTERED EXPERIMENTAL TRIALS

Response measure	Correlation statistic	Peak magnitude estimate	Simultaneous estimate difference	Overshoot duration	Test response duration
Peak magnitude estimates	Zero order	1.00***			
	Partial	1.00***			
Simultaneous estimate difference	Zero order	.84***	1.00***		
	Partial	.86***	1.00***		
Overshoot duration	Zero order	−.14**	−.03	1.00***	
	Partial	−.12*	−.05	1.00***	
Test response duration	Zero order	.40***	.35***	−.05	1.00***
	Partial	.39***	.37***	−.03	1.00***
Postreduction response duration	Zero order	.33***	.27***	.40***	.65***
	Partial	.35***	.25***	.38***	.69***

Note.—Partial correlation controlled for the effect of the length of *O*'s altered interval.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

between response measures, for which the duration of each *O*'s altered interval was the controlled variable. The similarity between the zero-order *r*s and partial *r*s indicate that the unique duration of altered intervals did not account for the significant *r*s between the various response measures. The similarity between the zero-order *r*s and partial *r*s is also consistent with the data in Table 3, which show that the length of each *O*'s altered interval was only weakly related to the response measures.

DISCUSSION

The torsion pendulum model of cupular functioning, based on anatomical and physiological characteristics of the semicircular canals, describes cupular movement as being heavily damped by the endolymph flow, which meets resistance within the narrow inside diameter of each canal (Guedry, 1965). This model has been used to design or interpret experiments investigating vestibular phenomena (Guedry, 1965). The design of the present study was based directly on the pattern of subjective response. No intervening model of cupular displacement was included because of the frequent systematic departures of subjective vestibular response from predictions of the torsion pendulum model and because of the great variation between subjective response patterns among *O*s during prolonged, constant *A*.

The method of basing the duration of the independent variable on the individual *O*'s subjective response pattern during a constant *A* appears to have been quite adequate. In spite of the fact that the *O*s' duration to peak subjective response during the base-line condition ranged from 18 to 51 sec., the pattern of responses to the various stimuli was consistent enough to yield a high level of significance for several important measures. The results of this experiment clearly show that both the magnitude of an altering *A* and the time in the adaptation cycle in which it was presented had significant effects on the size and pattern of the *O*s' magnitude estimates (Fig. 4 and 5). Furthermore, the data in Tables 3 and 4 show that the unique duration of the *O*s' altered interval had little systematic effect upon the various measures of subjective response.

The peak magnitude estimates for the 3° conditions were significantly greater than those for the 2° conditions. These results are in accord with the torsion pendulum model of cupular function and with the findings of other investigators (Brown, 1966, 1968; Clark & Stewart, 1968, 1970).

An analysis of the magnitude estimates following the four different times of presentation reveals that, except for Cond. D, the peak magnitude estimates became progressively greater as the *A* was altered later in the trial (Fig. 4 and 5, Table 1). The torsion pendulum model would predict the peak magnitude estimates for the first three intervals on the basis of cupular deflection. Cupular deflection was not instantaneous; thus the cupula had more time to move toward its maximum 1° position when the 1°/sec² *A* that preceded the increased *A* was of longer duration. The cupula was ultimately displaced to a more extreme position when the increased *A* was presented later in the trial because its position was initially more extreme when the altering *A* was introduced. This explanation fails to account for the decline of peak magnitude estimates for Cond. D. The decline, however, could be attributed to adaptation. The longer duration of 1°/sec² *A* that preceded Interval D allowed time for adaptation to develop and suppress the peak magnitude estimates. This explanation is supported by the fact that Interval D occurred while magnitude estimates declined during a prolonged constant *A* (Fig. 4 and 5).

The increase of simultaneous estimate differences from Interval C to Interval D in the 3° condition also supports the adaptation interpretation stated above (Table 1). The simultaneous estimate difference at least partially compensated for adaptation during Interval D since it measured the difference between the suppressed peak magnitude estimates and the corresponding suppressed magnitude estimates in the experimental, constant *A* condition. The fact that the simultaneous estimate differences decreased from Interval C to Interval D in the 2° condition suggests that this response measure did not entirely compensate for adaptation.

The profile of simultaneous estimate differences reported in Table 1 was similar to the profile of the peak magnitude estimates. The similar profiles are consistent with the clos relationship between the two measures show in Table 4. The significant interaction o simultaneous estimate differences between th size and position of increased *A* indicates th the position of the increased *A* had a grea effect in the 3° conditions than in the 2° conditions. The interaction between the same two variables for the peak magnitude estimates was not significant because the peak magnitude estimates were suppressed when the 3°/sec² *A* was presented during Interval D (Table 1).

The results in Table 2 show that the duration of overshoot for the 2° and 3° conditions were similar but that overshoot duration became progressively less as the *A* was altered later in the trial. The overshoot effect is obviously not in accordance with the torsion pendulum model, which involves a heavily damped system. If subjective response corresponded directly to cupular deflection, overshoot would mean that cupular deflection increased after reduction of the altering *A*. The heavily damped torsion pendulum system would preclude such an increase in cupular deflection. Similarly, response perseveration is an inadequate explanation since one would expect similar effects across subsequent altered intervals. Overshoot may be accounted for by a reaction latency, which became progressively less as the *A* was increased later in the trial. The decline could have been caused by adaptation that suppressed the increase in the size of the magnitude estimates following the reduction of the altered *A*. This explanation assumes that adaptation occurred during the early phases of subjective response, that is, while *O* was reporting increased subjective velocity.

The test response duration was important because it described the overall duration of subjective response to the same test stimulus for all conditions. The test response duration increased as the *A* was altered later in the trial (Table 2). In addition, the test response durations for Intervals B, C, and D were greater than the test response duration for the experimental, constant *A* condition. Cupular displacement and peripheral adaptation can account for these results. As the *A* was altered later in the trial, the cupula returned to its adapted 1° position at a somewhat later time during the subsequent *A*.

Peripheral adaptation, however, does not account for the lack of significant differences between the test response durations for the 2° and 3° conditions (Table 2). The torsion pendulum model would predict that the cupula would return to its 1° position at a later time when the altered *A* was 3°/sec² because of greater cupular displacement. The lack of significant differences between the two levels of stimulation may have been due to adaptation that is a function of the duration and magnitude of the preceding proximal stimulation.

Postreduction response duration was used to determine if response time in the altered conditions was a simple function of the time needed for the cupula to return from its maximum position to the position associated with a 1°/sec² prolonged *A* (Table 2). A peripheral adaptation model, such as that proposed by Hallpike and Hood (1953), would predict the similar postreduction response durations across the altered intervals. The receptors beyond the 1° position were stimulated for about the same amount of time for the various conditions. The 1°/sec² *A* that preceded the altered interval affected only those receptors associated with a 1°/sec² *A*, and these receptors had adapted before subjective response terminated in the altered conditions. Consequently, the postreduction response duration was a function of the receptors that were stimulated beyond the 1° position.

Previous investigators have shown that the size of the peak magnitude estimates for the negative aftereffect are related to the size of those for the positive effect (Brown, 1966; Clark & Stewart, 1968, 1970; Parsons, 1970). The nonsignificant effect that the manipulated variables had upon the size of the magnitude estimates for the negative aftereffect indicates that the relation was due to the degree of prior adaptation rather than to response stereotypy.

The results indicated that the peak magnitude estimates diminished significantly across subsequent trial blocks. The simplest explanation, supported by results from a previous investigation, is that habituation occurred, and subjective response systematically diminished as *O*s were exposed to subsequent prolonged stimuli (Brown, 1966). An alternative explanation is that the decline was an artifact of the magnitude estimate procedure. Stevens (1956) reported that *O*s remembered a weak standard auditory stimulus as being louder after giving ratio estimates of subsequent stimuli that were of moderate intensity. Since a very weak intensity level was used as the standard stimulus in the present study, *O* may have gradually raised the criterion level with which he identified his modulus. Consequently, the subsequent overall magnitude estimates would have gradually declined across trial blocks.

REFERENCES

Brown, J. H. Magnitude estimation of angular velocity during passive rotation. *Journal of Experimental Psychology*, 1966, **72**, 169–172.

Brown, J. H. Cross-modal estimation of angular velocity. *Perception & Psychophysics*, 1968, **3**, 115–117.

Clark, B., & Stewart, J. D. Magnitude estimates of rotational velocity during and following prolonged increasing, constant, and zero acceleration. *Journal of Experimental Psychology*, 1968, **78**, 329–339.

Clark, B., & Stewart, J. D. Effects of prolonged acceleration on perception of subsequent stimuli. Paper presented at the meeting of the Aerospace Medical Association, San Francisco, April 1970.

Edwards, A. L. *Experimental design in psychological research.* San Francisco: Holt, Rinehart & Winston, 1968.

Guedry, F. E. Psychophysiological studies of vestibular function. In W. D. Neff (Ed.), *Contributions to sensory physiology.* New York: Academic Press, 1965.

Hallpike, C. S., & Hood, J. D. Fatigue and adaptation of the cupular mechanism of the human horizontal semicircular canal: An experimental investigation. *Royal Society of London*, 1953, **41**, 542–561.

McNemar, Q. *Psychological statistics.* New York: Wiley, 1962.

Parsons, R. D. Magnitude estimates of the oculogyral illusion during and following angular acceleration. *Journal of Experimental Psychology*, 1970, **84**, 230–238.

Roberts, D. The effects of attention, knowledge of results, and paired association upon the levels of sensitivity of the semicircular canals. Unpublished master's thesis, San Jose State College, 1971.

Stevens, S. S. The direct estimation of sensory magnitudes—loudness. *American Journal of Psychology*, 1956, **69**, 1–25.

(Received January 11, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 273–279

INFLUENCE OF A VISUAL FRAME AND VERTICAL–HORIZONTAL ILLUSION ON SHAPE AND SIZE PERCEPTION[1]

ROBERT L. HOUCK, ROY B. MEFFERD, JR.,[2] AND GLENDA J. GREENSTEIN

Psychiatric and Psychosomatic Research Laboratory, Veterans Administration Hospital, Houston, Texas

A variety of visual frames and of stimulus shapes are examined in a series of three experiments to investigate the magnitude of the frame's effect on apparent size and apparent shape. Although a slight frame effect is demonstrated under very special conditions, its importance to the perceptual experience is questioned. The predominating effect of the vertical–horizontal phenomena is also explored.

The vertical–horizontal illusion (V–H) was first reported by Fick (1851) when he observed that a square imposed on a dark background looked like a vertical rectangle. Künnapas (1955a, 1955b, 1957a, 1957b, 1959a, 1959b, 1963), using the influence of the illusion on apparent size, attributed the V–H illusion, at least in part, to a frame effect imposed on the visual field by the natural elliptical orbit of the eye. He reported that the apparent length of a vertical line viewed through an artificial elliptical frame decreased as the frame was changed from a horizontal (i.e., similar to the natural frame imposed by the eye) to a vertical orientation (Künnapas, 1959a). Additionally, a line imposed in a frame, either elliptical or rectangular, appeared larger as it filled more of the frame (Künnapas, 1957b), and the effect was noted on both the length and the width of the line (Künnapas, 1955b).

Although most investigations of the V–H illusion have dealt with its influence on the relative apparent *sizes* of vertical and horizontal lines as opposed to the original observations of *shape* distortions, the two phenomena seem to be related. The purpose of this article is to explore the relationship between alterations in both apparent size and apparent shape resulting from a combination of the V–H illusion and a variety of imposed visual frames.

EXPERIMENT I

If the frame formed as a result of the oval shape of the eye has an influence on the apparent size of both the length and width of the line (Künnapas, 1955b), then it should also influence the apparent shape of two-dimensional geometric figures. It was hypothesized that the appearance of a square as a rectangle with its "long" dimension vertical would be augmented when surrounded by a rectangular frame with its long dimension horizontal, but would be diminished by a rectangular frame with its long dimension vertical.

Method.—The *Os*, seven males and three females, were laboratory personnel who had participated in other perceptual experiments so that they were experienced in analyzing and reporting their visual percepts and in making perceptual judgments. However, they were all naive with respect to the hypothesis being tested here.

The stimuli were solid flat figures: a 14.3-cm. square (SQ), a 14.3 × 14.0 cm. rectangle (R), a circle (C) with a diameter of 14.3 cm., or an ellipse (E), whose long and short axes were 14.3 cm. and 14.0 cm., respectively. The rectangle and ellipse were almost symmetrical; the ratio of the short: long axes was .98. The frame was a skeletal rectangle 24 × 19 cm. (inside dimensions) with a perimeter 1 cm. wide. When the frame was used, the stimulus was centered within it and both were at the same distance from *O*. The green fluorescing stimuli illuminated with dim indirect ultraviolet light, were presented under extreme reduction conditions (negating the influence of the natural elliptical frame created by the orbit of the eye) 2.5 m. before the seated *O* in his frontoparallel plane. Regard was binocular and *O* was instructed to close his eyes between judgments. The rectangles and ellipses were presented with the long axis either

---

[1] Appreciation is expressed to Michael E. Matlock for his technical assistance in the completion of this experiment.

[2] Requests for reprints should be sent to Roy B. Mefferd, Jr., Director, Psychiatric and Psychosomatic Research Laboratory, Veterans Administration Hospital, 2002 Holcombe Boulevard, Houston, Texas 77031.

TABLE 1

JUDGED SHAPES UNDER THREE FRAME CONDITIONS

Vertical frame				No frame				Horizontal frame			
True shape	Reported shape			True shape	Reported shape			True shape	Reported shape		
Noncurved stimuli											
	RV	SQ	RH		RV	SQ	RH		RV	SQ	RH
RV	**70**	24	06	RV	**72**	28	00	RV	**54**	44	02
SQ	57	**41**	02	SQ	45	**52**	03	SQ	27	**63**	10
RH	34	63	**03**	RH	26	60	**14**	RH	14	69	**17**
$\bar{X}$	53.6	42.6	3.8	$\bar{X}$	47.6	46.6	5.8	$\bar{X}$	31.6	58.7	9.7
Curved stimuli											
	EV	C	EH		EV	C	EH		EV	C	EH
EV	**77**	19	04	EV	**72**	24	04	EV	**46**	47	07
C	44	**52**	04	C	21	**74**	05	C	13	**69**	18
EH	21	57	**22**	EH	07	42	**51**	EH	13	38	**49**
$\bar{X}$	47.4	42.6	10.0	$\bar{X}$	33.4	46.6	20.0	$\bar{X}$	24.0	51.3	24.7

Note.—Entries are presented as percentages; the numbers in bold face, running in left right diagonals, express the percentage veridical judgments. Abbreviations: R = rectangle; SQ = square; E = ellipse; C = circle; V = vertical; H = horizontal. The frame was rectangular.

vertical (RV, EV) or horizontal (RH, EH) and, when used, the long axis of the frame was either vertical (FV) or horizontal (FH).

Ten judgments were made by each *O* for each of the six stimuli (i.e., RV, RH, SQ, EV, EH, C) in each of the three frame conditions (no frame, FV, FH). The viewing session was in two parts (90 judgments each) separated by a 15-min. rest period. Half of the *O*s received the curved figures in the first session and the noncurved figures in the second, and vice versa for the other half. The order of presentation was randomized across all conditions.

The *O* was told that he would be shown a series of squares, circles, rectangles, and ellipses one at a time with or without a frame, and that his task was to judge rapidly the shape of each figure, and if he judged it to be nonsymmetrical, whether its long axis was vertical or horizontal. All judgments were made in less than 5 sec. with a 2-sec. interstimulus interval.

Results and discussion.—The operation of the V–H illusion, resulting in overestimation of vertical extents, was expected to influence percepts in a systematic series: largely veridical percepts with a vertical rectangle or ellipse, fewer veridical percepts with a circle or square, and fewest with a horizontal rectangle or ellipse. The imposition of a frame was expected to modify this V–H effect in the direction reported by Künnapas (1955b), i.e., to increase the apparent extent of the dimension (vertical or horizontal) of the stimulus that most nearly filled the frame.

The data are presented in Table 1, where each matrix represents one frame condition. The *O*s' judgments for each condition are arranged in rows according to the veridical shape of the stimulus being judged and in columns according to *O*s' judged shape (percept) of the stimulus. The operation of the V–H illusion is seen as a shift in a row toward the left of the veridical diagonal in each matrix. For example, in the upper matrices, a vertical rectangle was most often reported veridically regardless of the frame, as is consistent with the expectations that the V–H illusion would exaggerate the vertical dimension. In the No-Frame condition the square was reported nonveridically to be a vertical rectangle almost half of the time (45%), also in accordance with the expected V–H effect. Without a frame, a horizontal rectangle appeared to be a square most of the time (60%), and it even appeared to be a vertical rectangle in 26% of the reports.

The degree to which the imposition of a frame around the stimulus influenced the V–H illusion is seen as shifts in the column totals from their base-line positions in the No-Frame matrices. The expected effect (Künnapas, 1957b) was that a vertical frame would reduce the effects of the V–H illusion, since the stimulus was much nearer the vertical than the horizontal sides of the frame. This result would be seen as a shift in the column totals of the table toward the right. The converse was expected when the long dimension of the frame was horizontal—the judged shapes should shift toward the left. Actually, the shifts were exactly opposite to the predicted directions. In both cases, the orientation of the frame was the critical factor: A vertical frame enhanced the V–H effect (i.e., by making vertical extents appear greater) while a horizontal frame diminished it.

Considering only the 18 diagonal cells of Table 1, where *O*s veridically judged the stimulus, both the interaction effects of Frame × Stimulus, $F\ (4, 36) = 3.82$, and Stimulus × Shape, $F\ (2, 18) = 6.19$, both $ps < .05$, were significant. Therefore, the simple effects contributing to the interactions were examined with the critical difference test (Winer, 1962). The Stimulus × Shape interaction resulted from a significant difference between the curved and rectangular figures for the horizontal ($p < .01$) and for the symmetrical ($p < .05$) stimuli. However, the V–H illusion operated to exaggerate the vertical extent and to reduce the apparent width of the vertical stimuli ensuring their veridical perception. This difference due to curvature extended to the nonveridical judgments; almost twice as many of these judgments were in the direction of the V–H illusion with the noncurved as with the curved stimili. This relative failure for *O*s to report V–H induced apparent shape distortions in curved figures suggests that the basic perceptual processes involved with these may be different from those of rectangular figures. This corresponds with reports by Hubel and Wiesel (1963, 1965) that throughout their extensive microelectrode recordings of single neurons in the visual cortex they failed to observe any cells which discharged preferentially to curved lines in the same way that they did to straight edges and angles. Possibly the perception of curved figures involves more of an inductive analytical process than does that of rectangular figures where the visual architecture appears to provide a means for a more "direct" perception.

With respect to the hypothesis under consideration the analysis of the simple effects contributing to the Frame × Stimulus interaction revealed that the vertical stimulus was judged veridically significantly ($p < .01$) more often than the symmetric stimuli (circle and square) in the vertical frame condition. This did not occur in the No-Frame and Horizontal Frame conditions. The horizontal stimuli were judged veridically less often ($p < .01$) than the other two stimuli classes in all frame conditions. These data, considered as a whole, fail to confirm predictions concerning the influence of the visual frame on apparent shape based on earlier findings of its influence on apparent size.

EXPERIMENT II

The methods used in Exp. I, in which *O*s reported the stimuli as having one of three discrete shapes, may have obscured a frame effect that would have been discernible had the judgments been of apparent size rather than of apparent shape. Three additional differences between the methods used in Exp. I and those of Künnapas's (1955b) original investigations of the frame effect could also have contributed to the failure to extend his findings: (*a*) the type of stimulus placed within the frame, (*b*) its location within the frame, and (*c*) binocular as opposed to monocular regard. The purpose of the second experiment was to explore these possibilities.

Method.—The *O*s, nine males and four females, were experienced in making perceptual judgments and had served in perception experiments previously; six were *O*s in Exp. I. The standard stimuli were a solid square (8.0-cm. sides), a solid circle (8.0-cm. diameter), a straight .35 cm. × 8.0 cm. bar, and an L formed from two such bars. The bars were presented in both vertical and horizontal orientation. Two frames in the same plane as the stimulus were used, a skeletal rectangle (24 cm. × 19 cm., inside dimensions) or an ellipse of the same dimensions, both with a perimeter width of .5 cm. The rectangular frame was used with the square and circle to maintain a similar condition with Exp. I, while the elliptical frame was used with the bar and L stimuli to approximate more nearly the stimulus conditions employed by Künnapas (1959a). The long dimension of the frame was either vertical or horizontal. The position of the stimulus within the frame was either centered or offset as described in the specific cases to follow. Two .35-cm.-wide comparison rods—one vertical and 30 cm. to *O*'s left and on the same plane as the standard stimulus, and the other horizontal and 30 cm. directly below the standard—were adjusted in length remotely by *O* to match the apparent lengths of the various standard stimuli. This adjustment was made by means of a hand control with a variable ratio of movement-to-comparison size. The standard stimulus was at *O*'s eye level, 2.5 m. before him in his frontoparallel plane. Extreme reduction conditions were used with the stimuli glowing green and sharply delineated under dim indirect ultraviolet light. Regard was monocular with *O*'s preferred eye. The *E* recorded the judgments from behind a black screen and adjusted the comparison rods between pairs of judgments alternately either to 20-cm. or 0-cm. length. As each stimulus was presented, *O* adjusted first the vertical comparison rod (which then was left in place) and then the horizontal comparison rod to the apparent size of one dimension of the standard stimulus (e.g., its height); then *E* recorded the judgments, reset the comparisons, and *O* adjusted each comparison rod to the other dimension of the standard. The apparent width of the single bars was not judged. Thus, four judgments were made for each stimulus under each condition, except for the single bars where only two judgments were made.

Results and discussion.—The data were separated into five groups to examine specific questions. Both the square and the circle were contrasted under the three frame conditions (i.e., no frame, vertical rectangular frame, and horizontal rectangular frame). The data were analyzed separately for judgments with the vertical and the horizontal comparison stimuli. None of the interactions were significant and, as in Exp. I, the frame had no significant effect in any condition. However, both analyses revealed a significant main effect of stimulus shape, vertical comparison, F (1, 12) = 16.43; horizontal comparison, F (1, 12) = 21.06, both ps $< .01$. In both, the dimensions of the square were judged to be larger than those of the circle. When the vertical comparison was used the vertical dimension of the figure was judged larger than the horizontal dimension, viz., 7.5 cm. vs. 7.22 cm., F (1, 12) = 5.33, $p < .05$. The effect was in the same direction with the horizontal comparison (vertical dimension = 8.33 cm. and the horizontal dimension = 8.55 cm.) although in this case the effect was not statistically significant. The significant effect is consistent with the view that the V–H illusion acted to magnify the apparent size of the vertical dimension relative to that of the horizontal.

The role of stimulus complexity was examined using three classes of stimuli already described: the single bar, either vertical or horizontal, the L-shaped stimulus, and the square. None of the interactions were significant, and neither was the main effect for complexity of stimulus. However, the main effects of the orientation of the standard and the orientation of the comparison were significant, F (1, 12) = 5.24, $p < .05$; and F (1, 12) = 48.22, $p < .01$, respectively. The vertical standard was judged larger than the horizontal standard (7.6 cm. vs. 7.2 cm. with the vertical comparison, and 8.7 cm. vs. 8.5 cm. with the horizontal comparison). This is consistent with the V–H illusion, which exerted a similar effect on the comparison line by increasing the apparent length of the vertical comparison as shown by its relatively smaller adjustment than the horizontal comparison when both were adjusted to the same stimulus. These results indicate that the V–H illusion exerts its effect independent of stimulus complexity.

The vertical bar and the L stimulus discussed earlier were tested in three frame conditions (viz., no frame, the vertical elliptical frame, and the horizontal elliptical frame). Only the vertical component of the L stimulus served as the standard. The stimuli were positioned within the frame so that the midpoint of the vertical line was at the center of the frame. Neither the interactions nor the main effects of the type of stimulus or of the frame condition were significant. However, the main effect of comparison line was significant

F (1, 12) = 49.15, $p < .01$; the vertical comparison was smaller (7.79 cm.) than the horizontal comparison (8.93 cm.) when both were adjusted to the same stimulus. This was expected from the operation of the V–H illusion. Again no effect of visual frame on apparent size was observed even though the stimuli closely approximated those used by Künnapas (1957b) in his original investigations.

The vertical bar and the L-shaped standard stimuli were used to examine the effect of position within the frame on apparent size. These stimuli were presented within a vertical and a horizontal elliptical frame. The stimulus was positioned with either the bottom of the vertical component of the stimulus at the center of the frame, or with the midpoint of the vertical component positioned there.

The vertical comparison line was adjusted significantly, F (1, 12) = 106.92, $p < .01$, smaller (7.8 cm.) than the horizontal comparison (8.9 cm.) when both were adjusted to the same standard stimulus. No other main effect was significant. Once again no influence on the V–H illusion was observed to result from changes in the visual frame in which the stimulus was viewed.

The L-shaped stimulus described earlier was placed in both the horizontal and vertical elliptical frames in three positions relative to the center of the frame—midpoint of the vertical bar, midpoint of the horizontal bar, and outer point at the intersection of the two bars. The orientation of the L was always one bar vertical and the other horizontal.

The interactions were not significant, but the main effects of the orientation (i.e., vertical or horizontal) of the comparison stimulus, F (1, 12) = 118.57, and of the orientation of the standard stimulus being judged, F (1, 12) = 11.6, both $ps < .01$, (i.e., the vertical or horizontal part) were significant. As expected, the V–H illusion resulted in the vertical comparison being adjusted smaller than the horizontal one (7.66 cm. vs. 8.88 cm., respectively). The V–H illusion also exerted an effect, again as expected, on the apparent size of the standard. The vertical standard appeared larger than the horizontal standard (8.38 cm. vs. 8.16 cm., respectively). The position within the frame did not cause significant changes in the V–H effect.

Considered together the results of Exp. II indicated that over a wide variety of stimulus types (geometric figures, a straight line, and the L-shaped stimulus) no effect due to a visual frame could be demonstrated by the use of *external* comparison lines. This was true even for stimuli and frames comparable to those used by Künnapas (1959a) in his investigation of these phenomena.

EXPERIMENT III

In view of this failure to replicate Künnapas' results under conditions similar to those he used, we attempted a more direct replication of one of his reports (Künnapas, 1957b). The one remaining difference between the present series of experiments and those of Künnapas is that he placed both the standard stimulus (the horizontal component of the L-shaped stimulus) and the comparison stimulus (the vertical component of the same stimulus) within the same frame. According to his observations, a nonsymmetrical frame would act to increase the apparent size of that component which is in closer proximity and to decrease the apparent size of the component which is farther from the frame. The overall result would be a magnification of any illusory effect exerted by the frame on any single component. In the series of experiments reported above a comparison line was used that was external to the frame and, therefore, supposedly not influenced by it. In this experiment both an external and internal comparison stimulus were used.

Method.—Twelve *O*s, eight males and four females, nine of whom had served in either one or both of the earlier experiments, were all experienced at reporting and analyzing their visual percepts.

The standard and comparison stimuli were bars .35 cm. wide. The fixed standard, 10 cm. long, was horizontal while the two adjustable comparison rods were vertical, the internal one intersecting the left end of the horizontal standard at a 90° angle to form an L, and the external one 30 cm. to the

TABLE 2

MEAN LENGTHS OF JUDGMENTS OF THE LENGTH OF A HORIZONTAL STANDARD BAR WITHIN A FRAME USING A VERTICAL COMPARISON BAR PLACED EITHER WITHIN OR OUTSIDE THE FRAME

Frame	Mean length of the internal comparison judgments (in cm.)			Mean length of the external comparison judgments (in cm.)
	Ascending	Descending	Combined	Combined
Vertical ellipse	9.39 (.40)	9.43 (.51)	9.41 (.45)	9.70 (.29)
Circle	9.20 (.44)	9.24 (.48)	9.22 (.45)	9.70 (.28)
Horizontal ellipse	9.17 (.46)	9.22 (.52)	9.20 (.49)	9.93 (.22)
Vertical rectangle	9.34 (.54)	9.39 (.57)	9.33 (.60)	9.73 (.40)
Square	9.20 (.46)	9.32 (.51)	9.26 (.48)	9.84 (.36)
Horizontal rectangle	9.11 (.49)	9.25 (.52)	9.18 (.50)	9.72 (.42)

Note.—Standard deviations are given in parentheses.

left of the internal comparison rod. Both comparisons were remotely adjustable by *O* using the same manual adjustment apparatus described in Exp. II. Six visual frames were used: a rectangle or ellipse 40 cm. × 24 cm. presented either vertically or horizontally and a square or circle 40 cm. in diameter. The point of intersection of the two internal stimuli was at the center of the frame with the horizontal bar extending to *O*'s right. The center of the frame was presented at *O*'s eye level, and in his frontoparallel plane at the same depth (2 m. from *O*) as the standard and comparison stimuli. The stimuli fluoresced green under dim indirect ultraviolet illumination, and they appeared sharp and clearly delineated in an otherwise dark field. The *E* sat behind a black curtain where he recorded *O*'s judgments and reset the apparatus between trials. Regard was monocular with *O*'s preferred eye, and *O*'s head was in a fixed head rest. Each frame condition was tested in separate testing sessions; therefore, each *O* was tested in six separate sessions over a period of not less than 1 wk. and not more than 2 wk. A random order for presentation of the frames was established, and then the starting point of this order was systematically varied across the 12 *O*s (i.e., two *O*s received each possible frame order). Each *O* made 32 adjustments of the internal comparison and 16 adjustments of the external comparison for each of the six frame conditions. Half of the judgments were ascending and half descending. For the first 16 judgments *O* was instructed to adjust only the vertical line inside the frame so that it appeared to be the same size as the horizontal bar. Starting with Judgment 17 and extending through Judgment 32, *O* was instructed also to adjust the external comparison so that it appeared equal to the horizontal bar after he had made his judgment with the internal comparison. The latter was left exposed until the second judgment was completed.

Results and discussion.—The results are presented in Table 2. None of the interactions or the main effect of frame shape were significant. However, both the main effects of the direction of the judgment (i.e., ascending vs. descending) and of frame orientation were significant, F (2, 22) = 4.10, and F (1, 11) = 5.37, both ps < .05, respectively. Ascending judgments were smaller than descending ones. Relative to the frame orientation these results confirm Künnapas's (1957b) report that as the orientation of the frame changes from vertical to symmetrical to horizontal, the apparent length of the horizontal line presented within the frame decreases when measured by an intersecting vertical comparison also within the frame. When the vertical comparison was outside the frame, there were no significant effects.

Wallach and McKenna (1960) demonstrated that locating the comparison stimulus in relative isolation rendered perception of its size indeterminant in the absence of distance cues. In the current experiment it might be suggested that the factor determining the amount of adjustment made with the external comparison is something nonvisual (e.g., the average amount of manual turning of the apparatus control). However, two factors present in the stimulus conditions of the current experiment render this explanation untenable. First, the external comparison was not isolated; rather, it was the same distance from the left side of the frame as was the internal comparison. Thus, the frame served as a common reference point for both comparison stimuli. Second, the adjustment apparatus had a variable ratio of movement to stimulus size assuring that *O* could not make reliable adjustments on the basis of nonvisual, mechanical cues. The present findings suggest that the two measuring techniques, external versus internal comparison, do yield different results. The asymmetrical frame exerts an effect on the internal vertical comparison opposite to but complimentary with its effect on the horizontal standard. This causes an increase in the overall effect of the frame so that the effect of magnitude is greater than the standard error of the measuring instrument. The frame's effect on the horizontal standard, as measured with the external comparison, is so slight as to be less than the small error inherent in the instrument, and, therefore, it is undetectable.

CONCLUSIONS

Collectively these results indicate that for the variety of visual frames investigated, although a visual frame does exert an influence on extents bounded by the frame as was shown by Künnapas's (1959a) extremely careful and detailed work, this influence is exceedingly slight. Only when both the standard and the comparison stimuli were inside the frame, where both were influenced simultaneously in opposite directions by the asymmetrical frame, was the effect discernible (Exp. III). When the comparison was external but still closely adjacent to the frame, no frame effect was evident. Further, in Exp. II, it was shown via an external comparison that neither the position of the standard within the frame nor its orientation within the frame influenced *O*'s judgments significantly. At the same time, however, the powerful effect of the V–H illusion was evident in all of the various conditions whether measured by apparent shape (Exp. I) or apparent size (Exp. II). Thus, to attribute any but the most minimal effect of the natural visual frame resulting from the elliptical orbit of the eye to the V–H illusion of similar visual phenomena is not supported.

REFERENCES

FICK, A. *De errore quodam optico asymenetria bulbi effecto.* Marburg: J. A. Kochii, 1851.

HUBEL, D. H., & WIESEL, T. N. Shape and arrangement of columns in cat's striate cortex. *Journal of Physiology*, 1963, **165**, 559–568.

HUBEL, D. H., & WIESEL, T. N. Receptive fields and functional architecture in two non-striate visual areas (18 and 19) in the cat. *Journal of Neurophysiology*, 1965, **28**, 229–289.

KÜNNAPAS, T. M. An analysis of the vertical-horizontal illusion. *Journal of Experimental Psychology*, 1955, **49**, 134–140. (a)

KÜNNAPAS, T. M. Influence of frame size on apparent length of a line. *Journal of Experimental Psychology*, 1955, **50**, 168–170. (b)

KÜNNAPAS, T. M. The vertical-horizontal illusion and the visual field. *Journal of Experimental Psychology*, 1957, **53**, 405–407. (a)

KÜNNAPAS, T. M. Vertical-horizontal illusion and surrounding field. *Acta Psychologica*, 1957, **13**, 35–42. (b)

KÜNNAPAS, T. M. The vertical-horizontal illusion in artificial visual fields. *The Journal of Psychology*, 1959, **47**, 41–48. (a)

KÜNNAPAS, T. M. Visual field and subjective center of a diamond. *The Journal of Psychology*, 1959, **47**, 305–316. (b)

KÜNNAPAS, T. M. The frame illusion: A rectification. *Perceptual and Motor Skills*, 1963, **17**, 369–370.

WALLACH, H., & MCKENNA, V. V. On size-perception in the absence of cues for distance. *American Journal of Psychology*, 1960, **73**, 458–460.

WINER, B. J. *Statistical principles in experimental design.* New York; McGraw-Hill, 1962.

(Received January 26, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 280–286

REWARD MAGNITUDE AND SEQUENCE OF MAGNITUDES AS DETERMINANTS OF RESISTANCE TO EXTINCTION IN HUMANS[1]

JOHN LAMBERTH[2] AND DENNIS G. DYCK
University of Oklahoma

Three experiments investigated magnitude of reward and resistance to extinction (Rn) using social reinforcers with human *Ss*. In Exp. I, a group that received small (S) reward magnitude showed more Rn than a group trained on large (L) magnitude of reward. In Exp. II, a group that had transitions from nonreward to large reward was more resistant to extinction than a group that had transitions from nonreward to small reward. In Exp. III, a group that experienced an S–L transition in acquisition showed more Rn than a group that received an L–S transition, even though magnitude of reward was identical for both groups. The results were interpreted in sequential terms.

Considerable interest has been evidenced in drawing analogs from animal conditioning to conditioning in humans. Much of this interest has centered on extensions of Hull-Spence theory (Lott & Lott, 1968; Weiss, 1968). Relatively little work has been done on the sequential theory of Capaldi (1967), probably because of its recent emergence. The studies involving sequential variables and human *S*s that have been reported in the literature have dealt primarily with the number of nonreinforced trials prior to a reinforced trial (N length) (Halpern & Poon, 1971; Lamberth, 1971; Litchfield & Duerfeldt, 1969; Meyers & Capaldi, 1970).

There are other variables besides N length of importance to sequential theory, and the present paper is an attempt to utilize one of them, magnitude of reinforcement, with human *S*s. Capaldi's (1967, 1970) sequential theory of instrumental learning is essentially a memory (aftereffects) model. "According to the present hypothesis, performance on some trial *Tn* is regulated extensively by the stimuli contributed from earlier or preceeding trials in the series [Capaldi, 1967, p. 67]." In essence, events generate distinctive stimuli that persist and are modified by subsequent events. The implications of this model for magnitude of reward and transitions between small and large reward in predicting subsequent extinction performance are discussed in conjunction with the three experiments reported here.

A corollary reason for the present studies has to do with current interest in instrumental conditioning utilizing social reinforcers with humans as *S*s (Byrne, 1971; Lamberth, 1971; Lamberth & Craig, 1970; Lott & Lott, 1968; Staats & Staats, 1958; Weiss, 1968). Few of these studies have investigated magnitude of reward (exceptions are Clore, 1966; Lamberth & Craig, 1970) partly because of the difficulty of quantifying social reinforcers and partly because it is difficult if not impossible to control drive level via any sort of deprivation. The problem of drive level has not yet been solved, but Byrne and Rhamey (1965) have successfully quantified the relationship between certain social stimuli. Partly because of these problems, but more importantly because of the assumption of sequential theory (cf. Capaldi, 1970) that extinction is more indicative of what an organism has learned than acquisition, the present investigations concentrate on the more sensitive extinction measure.

EXPERIMENT I

There is abundant evidence in the animal-learning literature that groups trained with small magnitude of reward in acquisition show more resistance to extinction (Rn) than groups trained on large magnitude of reward. Virtually every theory of Rn has an explanation of this phenomenon. Sequential theory is no exception. In general, sequential theory assumes that the extinction of an instrumental response is a process of stimulus generalization decrement. Each reward or series of rewards in acquisition occasions a distinctive internal stimulus that is conditioned to the instrumental response on a subsequent rewarded trial.

---

[1] This research was supported in part by a grant from the University of Oklahoma. The authors would like to express their appreciation to John Ezell of the College of Arts and Sciences.

[2] Requests for reprints should be sent to John Lamberth, Department of Psychology, University of Oklahoma, Norman, Oklahoma 73069.

In extinction, the internal stimulus is always nonreward (S^N). The *S*s conditioned to respond to a stimulus complex that closely approximates nonreward (such as small reward) should show greater Rn relative to a group that receives training to a stimulus further removed from nonreward, such as large reward. A group trained on small reward is expected, therefore, to show more Rn than a group trained on large reward.

Method.—The *S*s were 20 University of Oklahoma undergraduate students. Early in the semester they were pretested on a 60-item survey of attitudes (Byrne, 1971). Reinforcements were attitudes similar to each *S*'s own, eight per trial for large magnitude of reward and one per trial for small magnitude of reward.

The apparatus was a gray Masonite panel 76.2 × 76.2 cm., with a large lever in the center that could be depressed 38.1 cm. To the right of the lever was a ready light, which served as a signal to *S*s that a new trial had begun. To the left of the lever were seven buttons arranged in a semicircle, with the button on the left labeled "Unpleasant," the one on the right labeled "Pleasant," and the middle button labeled "Neutral." In the middle of the semicircle was a light, which came on when a button was pressed and remained on for 4 sec. To the right of the lever was a slit through which cards could be slipped. The lever was wired to two clocks. The first clock was activated as soon as the ready light came on and was terminated when *S* began to depress the lever. The second clock was then activated and was terminated when the lever reached the bottom of the channel through which it passed. When *S* entered the experimental room, he was told that the experiment concerned people's responses to certain social information. He was told that information might not always be available to him, but when it was, he was to signal *E* that he wanted that information by pulling the lever. The lever pulling was made to seem only incidental to the real purpose of the experiment. If information was available, it was received on a card through the slit in the apparatus. Whether *S* received a card or not, he was to indicate how he felt by pushing one of the seven buttons to the left of the lever. The *S* was informed that he could stop pulling the lever at any time he wished. If he stopped, the experiment was terminated. There were two groups with 10 *S*s in each group. Group S received small magnitude of reward and Group L received large magnitude of reward.

Each group received 7 trials of acquisition and 20 trials in extinction. Specifically, a trial began with the onset of the ready light to the right of the lever. The *S* then pulled the lever and received a reinforcement; following this, *S* pushed the button indicating how he felt. The light in the semicircle of buttons came on when *S* pushed a button, and remained on for 4 sec. The offset of the light under the buttons and the onset of the ready light signaling the start of the next trial were simultaneous. If *S* did not initiate a lever pull within 30 sec. after the start of a trial, the experiment was terminated. If this occurred in acquisition, *S*'s data were excluded from analysis. Two *S*s in the L group and one in the S group quit in acquisition. If it occurred in extinction, *S* was assigned a latency of 15 sec. on each clock for the remaining trials. This procedure was necessary because the results of pilot work and other investigations (Lamberth, Gouaux, & Davis, in press; Litchfield & Duerfeldt, 1969) indicated that if *S* is prompted to pull the lever by *E* after the start of the experiment, he pulls for many trials at a steady rate regardless of the reinforcement contingencies.

Results and discussion.—There were no terminal acquisition differences between the Groups, t (18) = 1.23. The extinction results are presented in Fig. 1. There is a clear superiority early in extinction of the Small-Reward group over the Large-Reward group. Over the first five blocks of trials, there is a between-groups difference, F (1, 18) = 4.003, $p < .06$.[3]

It is apparent that our *S*s performed as predicted. This preliminary exhibition of magnitude differences in extinction is encouraging, but obviously does not relate at all to the question of whether sequential theory is operative with human *S*s and

---

[3] In this experiment and Exp. III, we had no a priori way of knowing how many trials to run in extinction. Analysis of our data over all 10 blocks showed very marginal significance, F (1, 18) = 2.76, $p \leq .11$. The variability increased as extinction proceeded and obscured the differences. Of course, in a situation in which *S*s are receiving continuous reinforcement, they would be expected to extinguish rapidly. Therefore, we analyzed only the first 5 blocks of trials.

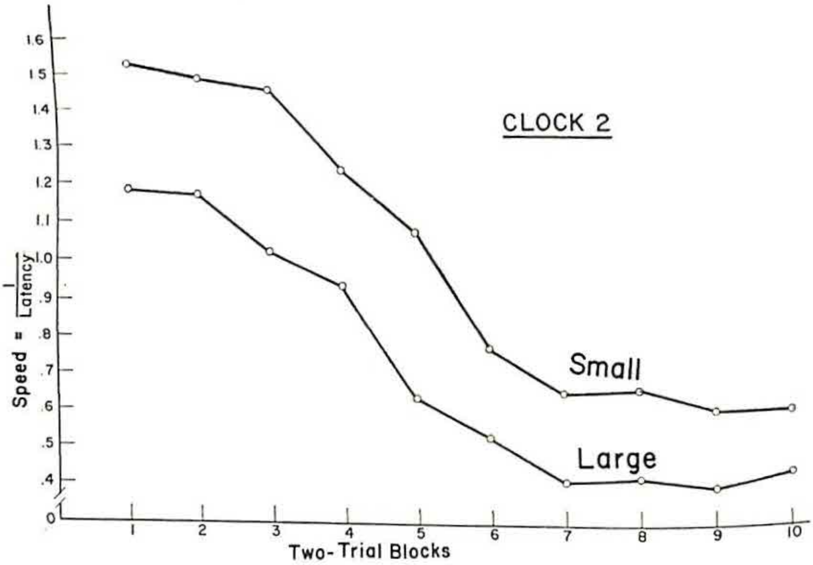


FIG. 1. Extinction curves for groups receiving small or large magnitude of reinforcement in Exp. I.

social reinforcers. The two experiments to follow were designed to make such a test.

EXPERIMENT II

Capaldi (1970) has argued that when shifts in reward magnitude and transitions from nonreward to reward are utilized, the variable of importance in determining Rn is the magnitude from which and to which *S*s are shifted. Capaldi assumes that nonreward (N) produces a distinctive stimulus (S^N), which occurs on the subsequent trial. If reward occurs on the next trial, S^N becomes conditioned to approach. The strength of this conditioning is stronger if the reward is large (L) rather than small (S). Conditioning is stronger for a group receiving N–L transitions than for a group receiving N–S transitions. The strength with which the instrumental response is conditioned to the stimulus trace of nonreward, in Capaldi's terms in S^N–R_I association, is the mechanism of the partial reinforcement extinction effect (PREE).

In the present experiment, rather than using merely the number of similar attitudes to differentiate magnitudes, we decided to use stimuli previously quantified by Byrne and Rhamey (1965). They report that personal evaluations are three times as powerful as attitudes in determining interpersonal attraction. Topics of low impact (analogous to our neutral statements) are half as effective in determining attraction as attitudes (Byrne, 1971). Furthermore, Lamberth and Craig (1970) have shown differences in performance in a free-responding choice situation when positive personal evaluations were used as large magnitude and "neutral statements" as small magnitude.

TABLE 1

ACQUISITION SCHEDULES FOR THE FOUR GROUPS IN EXPERIMENT II

Group	Trial																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
LNL	L	L	L	N	N	N	L	N	N	L	N	N	N	L	L	L	L	L	N	N	L	L
SNS	S	S	S	N	N	N	S	N	N	S	N	N	N	S	S	S	S	S	N	N	S	S
SNL	S	S	S	N	N	N	L	N	N	L	N	N	N	L	S	S	S	L	N	N	L	L
LNS	L	L	L	N	N	N	S	N	N	S	N	N	N	S	S	L	L	L	N	N	S	S

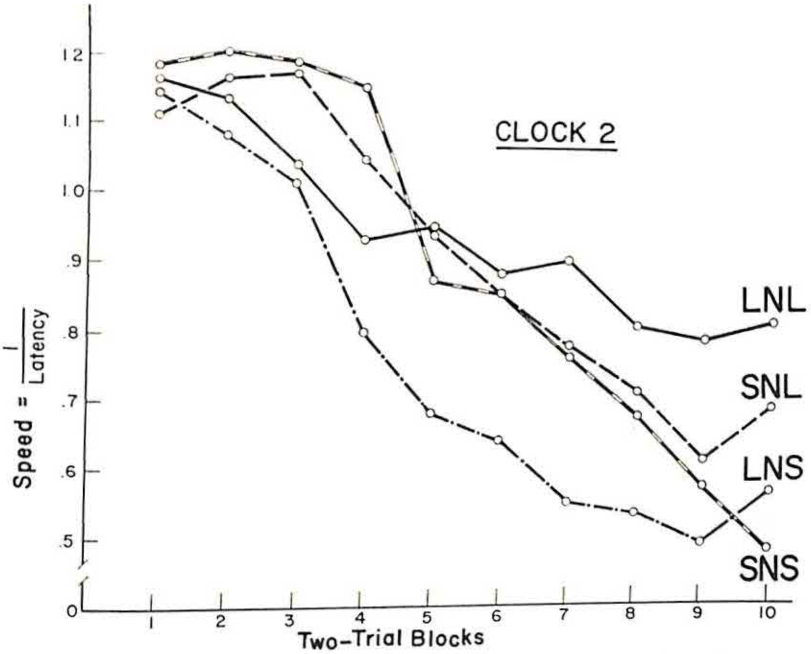


FIG. 2. Extinction curves for Groups SNS, SNL, LNS, and LNL in Exp. II.

Method.—The *Ss* were 64 University of Oklahoma undergraduate students. Early in the semester they were pretested on a 60-item survey of attitudes. These served as the basis of the personal evaluations. Each was told that another student had read his attitude survey and evaluated him on the Interpersonal Judgment Scale (IJS) (Byrne, 1961). The IJS consists of six 7-point rating scales concerning intelligence, morality, adjustment, knowledge of current events, how much the student would like *S*, and how much he would like to work with *S*. The scales range from very positive to very negative. Only the two most positive evaluations for each scale were used. For example, "I feel that I would probably like this person very much" and "I feel that I would probably like this person" were the only two statements for the liking scale which were used. On the basis of information gathered from a different sample, it was determined that *Ss* in the population felt that an intelligent premedical student was extremely competent. To increase the saliency of the evaluations, the person evaluating *S* was described as a premedical student with an IQ of above 135. The *S* was told that several students from the premedical club had evaluated him and that they had written their evaluations on cards. These evaluations were utilized as large magnitude of reward. Neutral statements such as "Red China has been in the news recently" served as small magnitude of reinforcement.

The apparatus was the same as the one used in Exp. I. The instructions were changed to explain that *S* might receive statements of fact or evaluations.

Four experimental groups were created with 16 *Ss* in each group. The first group started with small reward, was shifted to nonreward, and then back to small reward (SNS). The second began with small reward, was shifted to nonreward, and then to large reward (SNL). The third originated with large, was shifted to nonreward, and then to small reward (LNS); while the fourth started with large, was shifted to nonreward, and then back to large reward (LNL). The schedules, which were used by Capaldi (1970), are shown in Table 1. There were 22 trials of acquisition and 20 trials of extinction.

According to sequential theory, the critical variable is the magnitude of reward to which *Ss* are shifted following nonreward.[4] We hypothesized that Groups SNL and LNL would not differ from each other, but would show more Rn than Groups SNS and LNS, who also would not differ from each other.

Results and discussion.—The results are presented graphically in Fig. 2. Since we were in essence replicating a study reported by Capaldi (1970), we expected the same ordering of the groups as Capaldi reported. Capaldi reported that Groups LNL and SNL did not differ from each other but showed more Rn than Group SNS, which was faster than Group LNS. The groups

---

[4] Of course, we are ignoring the transitions from reward to nonreward (R–N transitions). The effects of these transitions are less robust than N–R transitions, and probably we would not expect them to have any effect with fairly small differences in magnitude and no control of drive level.

Preliminary evidence from these investigations and at least four others (Halpern & Poon, 1971; Lamberth, 1971; Litchfield & Duerfeldt, 1969; Meyers & Capaldi, 1970) indicate that researchers interested in human conditioning should consider these more specific operations when studying an analog to instrumental reward conditioning. The effects of sequence of reward in acquisition are, according to Capaldi, a fundamental difference between sequential and nonsequential theory. The sequence of reward is not viewed as just another variable, such as drive, but rather organisms which experience different sequences learn different things. This follows from the proposition that different reward events have distinctive internal stimuli, and organisms learn responses to different associations or cognitions with different training sequences. In many ways, with its emphasis on memory and learning cognitions, etc., sequential theory seems even more suitable to humans than to rats.

One further word should be said about the reinforcers used here. These are, of course, some of the reinforcers used in social situations. Perhaps it is premature, but it is intriguing to speculate about the possibility that there may be sequences of reinforcement (with social reinforcers) that would make interpersonal relationships maximally resistant to extinction.

REFERENCES

Amsel, A. Partial reinforcement effects on vigor and persistence. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 1. New York: Academic Press, 1967.

Byrne, D. Interpersonal attraction and attitude similarity. *Journal of Abnormal and Social Psychology*, 1961, **62**, 713–715.

Byrne, D. *The attraction paradigm.* New York: Academic Press, 1971.

Byrne, D., & Rhamey, R. Magnitude of positive and negative reinforcements as a determinant of attraction. *Journal of Personality and Social Psychology*, 1965, **2**, 884–889.

Capaldi, E. J. A sequential hypothesis of instrumental learning. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 1. New York: Academic Press, 1967.

Capaldi, E. J. An analysis of the role of reward and reward magnitude in instrumental learning. In J. Reynierse (Ed.), *Current issues in animal learning.* Lincoln: University of Nebraska Press, 1970.

Clore, G. L. Discrimination learning as a function of awareness and magnitude of attitudinal reinforcement. Unpublished doctoral dissertation, University of Texas, 1966.

Gonzales, R. C., & Bitterman, M. E. Spaced-trials partial reinforcement effect as a function of contrast. *Journal of Comparative and Physiological Psychology*, 1969, **67**, 94–103.

Halpern, J., & Poon, L. Human partial reinforcement extinction effects: An information-processing development from Capaldi's sequential theory. *Journal of Experimental Psychology*, 1971, **89**, 207–227.

Hull, C. L. *A behavior system.* New Haven: Yale University Press, 1952.

Jonckhere, A. R. A distribution-free k-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

Lamberth, J. The effect of sequential variables on performance using attitudinal reinforcers. *Psychonomic Science*, 1971, **22**, 350–351.

Lamberth, J., & Craig, L. Differential magnitude of reward and magnitude shifts using attitudinal stimuli. *Journal of Experimental Research in Personality*, 1970, **4**, 281–285.

Lamberth, J., Gouaux, C., & Davis, J. Agreeing attitudinal statements as positive reinforcers in instrumental conditioning. *Psychonomic Science*, in press.

Lawrence, D. H., & Festinger, L. *Deterrents and reinforcement—The psychology of insufficient reward.* Stanford: Stanford University Press, 1962.

Litchfield, P. M., & Duerfeldt, P. H. Resistance to extinction in children as a function of N-length and number of different N-lengths. *Psychonomic Science*, 1969, **14**, 299–300.

Lott, A. J., & Lott, B. E. A learning theory approach to interpersonal attitudes. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes.* New York: Academic Press, 1968.

Meyers, L. S., & Capaldi, E. J. Resistance to extinction as a function of sequence of events in partial reinforcement. *Psychonomic Science*, 1970, **19**, 199–200.

Staats, A. W., & Staats, C. K. Attitudes established by classical conditioning. *Journal of Abnormal and Social Psychology*, 1958, **57**, 37–40.

Weiss, R. F. An extension of Hullian learning theory to persuasive communication. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes.* New York: Academic Press, 1968.

(Received January 29, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 287–290

EFFECTS OF 24-HOUR SLEEP DEPRIVATION ON RATE OF DECREMENT IN A 10-MINUTE AUDITORY REACTION TIME TASK[1]

HANS-OLOF LISPER[2] AND ANDERS KJELLBERG

University of Uppsala, Uppsala, Sweden

A reaction time (RT) task of short duration, characterized by a high signal rate (16 signals/minute) was used to examine if 1 night of sleep deprivation (SD) in interaction with task duration did transfer and/or change the RT distribution. Eight *S*s participated on two occasions: after SD and after normal sleep. From a significant interaction between SD and task duration for 25th, 50th, and 75th percentiles, it was concluded that there was a general increase in RT. Strengthening of the interaction for the 75th percentile values and significantly more blockings for the last minutes in SD also did confirm an increased skewness found by other researchers. Since these effects are a function of interaction, it is suggested that one of the effects of SD can be an accentuation of the decrement found in vigilance.

One of the characteristics of tasks sensitive to sleep deprivation (SD) is that they have to be of a certain duration. Wilkinson (1968) summarizing the influence of test duration stated that a test will be sensitive to 30 hr. of SD only if it is prolonged for at least 30 min. This is a rather cautious recommendation as effects of SD have been demonstrated with shorter tests (e.g., Donnell, 1969; Williams, Granda, Jones, Lubin, & Armington, 1962); but even for these, time on task seems to be an important factor. That an SD test must be prolonged is dependent upon the fact that SD does not seem to affect the initial level of performance, but only the rate of decrement. Thus after SD the deterioration of performance is more rapid, and this will sooner or later result in a significant difference between the SD and the normal sleep condition. If increased rate of decrement is the pervading effect of SD, other tasks that show decrement over time under normal conditions ought to be sensitive to SD.

In the present experiment, a 10-min. task with this characteristic was used. This task was developed from the findings by Lisper (1969) that reaction time (RT) to 90-db. auditory signals increased over 3 hr. when using a high signal rate (240 signals/hr), but was unaffected when the signal rate was low (12 signals/hr). Lisper and Ericsson (1972) further increased the signal rate and found that with 15 signals/min, a 10-min. watch was sufficient for a significant increase in RT. In the reported experiment, a similar 10-min. task was used; i.e., RT will be measured to 70-db. signals presented at a signal rate of 16 signals/min.

That this task should be sensitive to SD is also supported by the findings of Williams, Lubin, and Goodnow (1959), who found that a 10-min. choice RT task was sensitive to SD. They also found that RT increased with task duration, especially if the task was without interruptions. When analyzing the result with SD as independent variable, they did not consider the interaction between SD and task duration and treated the 10-min. period in its entirety. Doing this, they demonstrated that the 10 longest RTs increased over hours of SD, but that on the other hand, the 10 shortest RTs were rather unaffected. This mode of analysis leaves open the possibility that the 10 shortest RTs were registered at the beginning and the 10 longest at the end of the 10-min. test. Thus it is possible that SD in interaction with task duration affects even the short RTs. In the reported experiment, difwere ranked as predicted, and the data were analyzed by the Jonckhere test (Jonckhere, 1954), which examines the overall difference between groups and their rank orderings in a single operation. The analysis indicated a marginally significant effect, $z = 1.32$, $p = .09$. The groups are ordered as expected and indicate that a transition from nonreward to large reward renders *S*s more resistant to extinction than a transition from nonreward to small reward.

---

[1] The investigation was supported by grants from the Swedish Council for Social Science Research.

[2] Requests for reprints should be sent to Hans-Olof Lisper, Department of Psychology, University of Uppsala, Svartbäcksgatan 10, S-753 20 Uppsala, Sweden.

Seemingly, sequential theory is the only major theory of instrumental learning that would hypothesize differences in Rn between groups which experience a transition from nonreward to large reward versus groups which experience a transition from nonreward to small reward. Other theories (Amsel, 1967; Lawrence & Festinger, 1962) assume that Rn is an increasing function of magnitude of partial reward; that is, as the magnitude of reward increases, dissonance or frustration increases and regulates the PREE. If their interpretation is correct, then Group LNL should show most Rn, with SNL and LNS next, and Group SNS least resistant. It is obvious from our results that these predictions did not obtain.[5]

Several critics of sequential theory have argued that sequential variables are only important in explaining the PREE when massed trial acquisition is used (Amsel, 1967; Gonzales & Bitterman, 1969; Hull, 1952). Obviously massed trials were used here in acquisition. Recently Capaldi (1970) has shown that the same schedules as employed here with a 24-hr. intertrial interval yield the same results as massed acquisition trials.

EXPERIMENT III

A sequential extension of the N–R transition concept explored in Exp. II is the idea that transitions in reward magnitude should also influence Rn. The logic of the idea is straightforward. In a transition from small to large (S–L) reward, the stimulus produced by small reward is conditioned to the instrumental response by means of large reward. In L–S transitions, on the other hand, the stimulus produced by a large reward is conditioned by means of a small reward. Approach is more strongly conditioned to the internal stimulus of small reward in Group S–L than in Group L–S. The stimulus occurring in extinction is more similar to the stimulus of small reward and thus Group S–L will have an approach response more strongly conditioned to a stimulus complex that approximates N than will Group L–S.

One complication arises when using humans as *S*s in a study of this type. Typically, experiments of this sort are run two trials a day with the appropriate transitions. It is assumed that the 24-hr. intertrial interval to the next day will weaken what is in effect the opposite type of transition. It is impractical to bring humans back to the lab over a period of time to allow many transitions. For this reason, only one transition was possible.

Method.—The *S*s were 39 University of Oklahoma undergraduates. The apparatus was the same as the one used in Exp. I. Following Capaldi (1970), three groups were run, 13 *S*s per group. Group S–L received 10 trials of acquisition reinforced by neutral statements followed by 10 trials reinforced by positive personal evaluations. Group L–S received positive evaluations for the first 10 trials and neutral statements for the final 10 trials. Group L–L received positive evaluations on all 20 trials. All *S*s were run in 20 trials of extinction.

Results.—The results are presented graphically in Fig. 3. Statistical analysis across all 10 blocks of trials indicated only a significant trials effect, F (9, 324) $= 16.72$, $p < .001$. It appeared, as in Exp. I, that *S*s extinguished rapidly. This, of course, is reasonable in that in Exp. I and III, *S*s were continuously reinforced and would be expected to extinguish more rapidly than *S*s receiving partial reinforcement in Exp. II.

Additional analysis confirmed the speculation that we had run too many extinction

---

[5] It would be extremely naive to assume that we are theory testing in the present situation. These learning theories have been developed using another species and other reinforcers. Any "test" they may face should rightly be conducted with rats and primary reinforcers. On the other hand, if the theory predicts behavior across species and reinforcers, generality has been added to that theory.

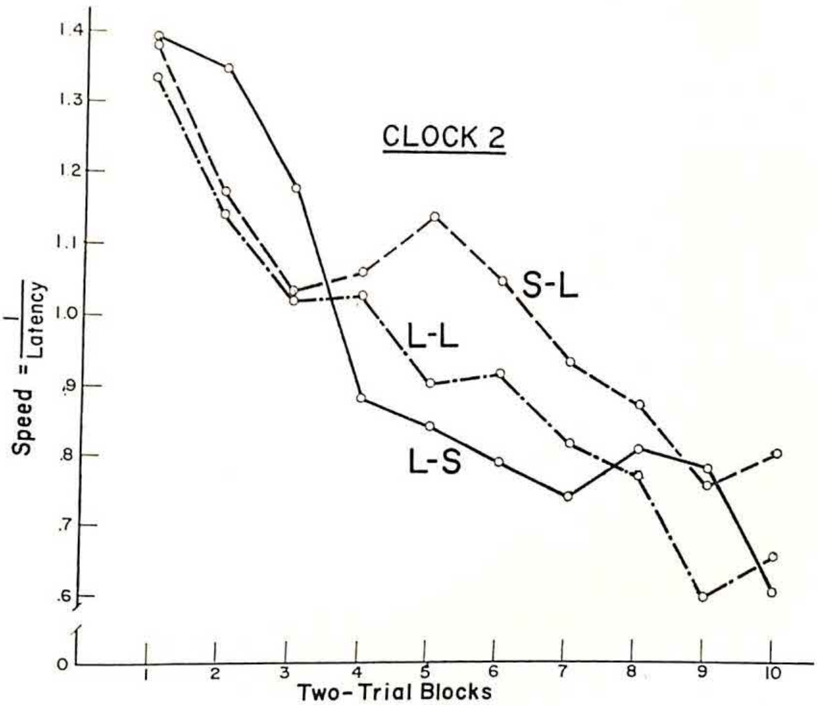


FIG. 3. Extinction curves for Groups LL, SL, and LS in Exp. III.

trials and, in effect, obscured real differences. Analysis of variance across the first five blocks of trials indicated a significant trials effect, F (4, 144) = 16.34, $p < .001$, and a significant Groups × Trials interaction, F (8, 144) = 2.25, $p < .03$. Post hoc analysis (Scheffé's) indicated that there were no reliable differences between groups early in extinction. In fact, Group L–S was faster across the first three blocks of trials than Group S–L. By Trial Block 5, Group S–L was significantly faster than Group L–S ($p < .01$), indicating that Group S–L was more resistant to extinction, as predicted. Group S–L and L–L did not differ from each other although Group L–L tended to show less Rn than Group S–L. This, of course, is consistent with sequential theory in that Group L–L had not experienced the stimulus most similar to N in acquisition (i.e., small reward).

GENERAL DISCUSSION

The present experiments lend support to the idea that sequential variables play a major role in determining Rn with human *S*s and social reinforcements. It is obvious that N–R transitions and S–L transitions are determinants of Rn in humans. This is the case even though in Exp. III there was only one S–L transition.

Virtually all the viable theories of extinction can predict the results we obtained in Exp. I. As was indicated earlier, only Capaldi's sequential theory can predict the results of Exp. II, and the same is true for Exp. III. Groups L–S and S–L had exactly the same number and magnitude of reinforcement in acquisition. Both groups should have experienced the same level of frustration or dissonance in extinction and thus been equally resistant to extinction. It seems apparent that what is important in determining Rn in a varied magnitude situation is not magnitude per se, but rather the sequence of small and large rewards.

The intriguing possibilities in generalizing via sequential principles from rats to humans are many. The most general point that Capaldi (1967, 1970) emphasizes is that it is necessary to analyze acquisition into more specific components than partial reward or varied magnitude of reward. To predict Rn accurately, schedules of reward need to be broken down more specifically into dyadic sequences (e.g., R–R, R–N, N–R, N–N).

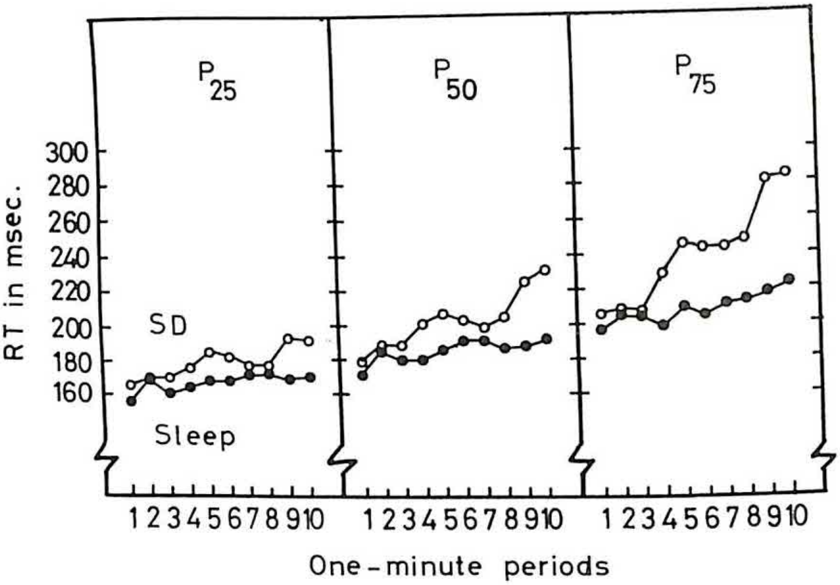


FIG. 1. Changes in 25th, 50th, and 75th percentile reaction times over 10 min. after normal sleep (closed circles) and sleep deprivation (open circles).

ferent parts of the RT distribution will be considered to investigate if 1 night of SD in interaction with task duration will transfer the entire distribution toward longer RTs or if only the long RTs are affected.

METHOD

Five female and three male undergraduate students, between 20 and 25 yr. of age, participated as paid *S*s.

The experiment was conducted in a 1.5 × 2.0 × 2.0 m. semisoundproof chamber, which was dimly lit during the experiment. The *S*s were comfortably seated in a half-reclining position in an armchair and were instructed to press a microswitch, held in the preferred hand, as rapidly as possible when they heard an auditory signal. This response terminated the signal, which came from a loudspeaker placed in front of and above *S*s. The signal was a 1,000-Hz. tone produced by an Oltronix RC-oscillator Type RCO-6K and having the intensity 70 db. re .0002 dyn/cm² measured at the place of *S*s' heads. The mean interstimulus interval (ISI) was 3.75 sec. with a variation of ±1.25 sec. Eleven intervals were used all with the same probability, and the order between them was randomized over the 10-min. watch. The *S*s' RTs were recorded in 1-msec. units by an electronic timer, manufactured by the Department of Psychology, University of Uppsala, Sweden, and were then printed on an ADDO-X printer.

Each *S* participated on two occasions with an interval of 6 to 8 days. Half of the *S*s started the experiment after 1 night of SD and the other four *S*s after a night of normal sleep. Two *S*s at a time were kept awake, and this was supervised from 10 p.m. until the test started the next morning between 8:00 and 8:30 a.m. In the SD condition, then, *S*s had been without sleep for about 24 hr. Before the test started, *S*s were to surrender their watches and were on both occasions given 1-min. practicing on the RT task.

RESULT

For each *S*, the 25th, 50th, and 75th percentile RTs were computed for each period of 1 min. Table 1 shows the summaries of analysis of variance (Edwards, 1960, pp. 224–253) based on these values. This analysis shows that there were significant linear trends over blocks of time for all three parts of the RT distribution. There was also a significant

TABLE 1

ANALYSIS OF VARIANCE OF 25TH, 50TH, AND 75TH PERCENTILE REACTION TIME

Source	*df*	F-value		
		25th	50th	75th
Time blocks (T)	9,63	4.97***	4.70***	4.50***
T linear trend	1,63	7.15**	37.45***	37.97***
Degree of SD (S)	1,7	6.37*	5.98*	4.44
B × S	9,63	1.46	1.62	1.52
B × S linear trend	1,63	4.87*	8.58**	11.29**

* $p < .05$.
** $p < .01$.
*** $p < .001$.

main effect of SD. More interesting was that SD interacted with the linear trend of blocks of time for all parts of the RT distribution. Figure 1 shows the nature of these effects. There was an increase of RT as a function of task duration in both conditions, but the increase was more pronounced after SD. From Fig. 1 and Table 2, (intercepts) it can be seen that the difference between the two conditions was negligible for the first minutes of the test. From Table 2 it can also be seen that there was a general displacement of the RT distribution upwards, but that the long RTs were most affected by time on task. Further it can be seen that the interaction between SD and blocks of time affected the entire RT distribution, but also that the effect was more pronounced for the long RTs. To get an idea of the necessity of a 10-min period of testing, the same analysis was done for the first 5 minutes of the test. The linear effect of blocks of time was, even in this case, significant, F (1, 63) = 7.15, $p < .01$, but the main effect of SD was not significant. The linear interaction, on the other hand, was still significant for 50th and 75th percentiles, F (1, 28) $\geq$ 4.72, $p < .05$.

Blockings, defined as RTs twice the period average in the sleep condition, were significantly increased by SD ($p = .011$. Sign test, Siegel, 1956, pp. 68–75). This difference was a function of blockings in the last 2 min. of the test.

DISCUSSION

The result showed that a short simple RT task was sensitive to 24 hr. of SD. The fact that the test is short has one important methodological advantage; it is possible to make repeated measurements within the same SD session, without having the result too much blurred by, e.g., circadian rhythms and/or progressive effects on motivation. The significant effect of SD during the first 5 min. of the task suggests that the task even could be considerably shortened. This would further diminish the effects of extraneous factors.

From the reported experiment it is concluded that even the short RTs are affected by SD. Williams et al. (1959) on the other hand, found that the 10 shortest RTs were affected very little by even rather extended periods of SD. Since Williams et al. (1959) did consider the 10-min. period in its entirety these two conclusions might not be contradictory. Thus the present authors, in retrospect, would predict that the fastest responses found by Williams et al. (1959) were scored at the first few minutes of the test, and thus their conclusion might be another way of saying that the *initial* performance is unaffected by even very long periods of SD.

Wilkinson (1968) concluded that an SD test has to be prolonged for at least one-half hour to be sensitive to moderate SD. Is then the present task more sensitive to SD than, e.g., Wilkinson's (1961) five-choice test of serial reaction? The answer to this problem is not only tied to the effects of SD per se, but also to the measures and statistical analysis used. Concerning the reported task, RT ought to be a more sensitive measure of decrement than, e.g., the detection measures used in vigilance. According to Buck (1966), there is a deterioration in RT even before detections are affected. That in an RT task every signal is responded to in a graded fashion, and not in the less informative yes/no way of detection measures, means further statistical gains. In the analysis of the reported experiment, the rate of decrement was of main concern and consequently an analysis of trend was considered most relevant. That this is a more powerful statistical test for this aim than the ordinary analysis of variance was shown by the fact that the most important interaction between SD and time on task was significant only if analysis of trend was used. Thus a comparison between two different tests is very complicated and Wilkinson (1965) himself also stated that only broad comparisons could be made. Consequently the problem of the sensitivity of the task must be left without an answer. Furthermore the present authors would like to stress that research must be concentrated on the mechanisms behind the effects of SD and not

TABLE 2

INCREASE AND INTERCEPTS (IN MSEC.) OF STRAIGHT LINES FITTED TO REACTION TIME DATA

Percentile	Normal sleep		Sleep deprivation	
	Increase	Intercept	Increase	Intercept
25th	11	161	24	166
50th	16	176	44	179
75th	22	197	80	196

let it suffice with categorizing tasks into sensitive and insensitive ones.

Concerning mechanisms, Williams et al. (1959) stated that the deficit from SD on all their tasks "took the form of lapses [brief periods of no response accompanied by extreme drowsiness and a decline in EEG alpha amplitude, p. 24]." They also mentioned that there might be a gradual deterioration affecting every RT. The reported experiment showed explicitly that SD in interaction with time on task gave a deterioration affecting the shorter RTs even when the period of SD was as short as 24 hr. The result also confirmed the increase in positive skewness and number of blockings found by Williams et al. (1959). This increase was a result of SD in interaction with time on task. Thus for the first 3 min., the RT and number of blockings in the two conditions were very similar. If increased skewness and number of blockings should be interpreted as a function of lapses (instances of lowered arousal) and/or in line with Broadbent's (1958) filter theory (attentional concepts), it is difficult to judge from the reported experiment without adding other measures, preferably EEG.

In the search for mechanisms behind the effects of SD, another point of departure could be the fact that rate of decrement is accentuated by SD. Thus one of the mechanisms behind the effects of SD might be as simple as adding a constant to a more basic process. Consequently the present authors would like to strongly stress that SD cannot be considered in isolation and that theories of vigilance might deal with the suggested basic process. Of these theories, the habituation theory by Mackworth (1968) seems to fit best into data from tasks very similar to those reported, in non-SD conditions. Thus higher signal rates gave faster decrement than low rates (Lisper, 1969; Lisper & Ericsson, 1972). Further the extension of the habituation theory to SD is also favored by Mackworth herself (1969, pp. 132–144).

REFERENCES

Broadbent, D. E. *Perception and communication.* London: Pergamon Press, 1958.

Buck, L. Reaction time as a measure of perceptual vigilance. *Psychological Bulletin*, 1966, **65**, 291–304.

Donnell, J. M. Performance decrement as a function of total sleep loss and task duration. *Perceptual and Motor Skills*, 1969, **29**, 711–714.

Edwards, A. L. *Experimental designs in psychological research.* (Rev. ed.) New York: Holt, Rinehart & Winston, 1960.

Lisper, H.-O. Classical reaction time and signal rate in a vigilance setting. *Psychonomic Science*, 1969, **17**, 217–218.

Lisper, H.-O., & Ericsson, S. *Signaltakt och RT-ökning över tid* (Preliminary report) Uppsala, Sweden: University of Uppsala, 1972.

Mackworth, J. F. Vigilance, arousal, and habituation. *Psychological Review*, 1968, **75**, 308–322.

Mackworth, J. F. *Vigilance and habituation.* Harmondsworth: Penguin Books, 1969.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Wilkinson, R. T. Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. *Journal of Experimental Psychology*, 1961, **62**, 263–271.

Wilkinson R. T. Sleep deprivation. In O. G. Edholm & A. L. Bacharach (Eds.). *The physiology of human survival.* New York: Academic Press, 1965.

Wilkinson, R. T. Sleep deprivation: Performance tests for partial and selective sleep deprivation. In L. A. Abt & B. F. Riess (Eds.), *Progress in clinical psychology.* Vol. 8. New York: Grune & Stratton, 1968.

Williams, H. L., Granda, A. M., Jones, R. C., Lubin, A., & Armington, J. C. EEG frequency and finger pulse volume as predictors of reaction time during sleep loss. *Electroencephalography and Clinical Neurophysiology.* 1962, **14**, 64–70.

Williams, H. L., Lubin, A., & Goodnow, J. J. Impaired performance with acute sleep loss. *Psychological Monographs.* 1959, **73**, (14, Whole No. 484).

(Received February 1, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 291–296

MEANINGFULNESS AND PRONOUNCEABILITY AS CHUNKING UNITS IN SHORT-TERM MEMORY[1]

DONALD W. McMURRAY[2] AND THOMAS M. DUFFY

University of Manitoba, Winnipeg, Manitoba, Canada

The experiments distinguished the effects of meaningfulness (M) and pronounceability (P) on acquisition and retention. In the first experiment, the exposure duration was varied and immediate recall obtained. It was found that P strings required a shorter exposure duration than M strings to obtain the same recall level. Random (R) strings required the longest exposure duration. In the second experiment, level of learning was equated by the use of the previously determined exposure times and a retention interval was introduced. Delayed recall was highest for R strings and lowest for P strings. The results were interpreted in terms of differential representation of the chunks in short- and long-term memory due to the nature of the chunks and the opportunity for rehearsal.

The facilitating effect of grouping or chunking operations on recall in the short-term memory (STM) paradigm has been demonstrated with a variety of materials (e.g., Bower & Winzenz, 1970; Laughery & Pinkus, 1968; Miller, 1956). However, almost invariably, the research has tested only immediate recall. Thus the separate effects of chunking operations on acquisition and retention are impossible to disentangle. The focus of the present research is to determine these separate effects for meaningful (M) and pronounceable (P) chunking units. Retention, here, will be considered the level of recall after a retention interval when level of immediate recall has been equated. The retention interval must involve presentation of a task which will minimize rehearsal opportunity for *S*. Acquisition is taken to be reflected in the level of immediate recall. An alternative acquisition measure is the exposure duration of a stimulus needed for recognition or identification.

Initial research on P and M strings as chunking units had generally indicated that while both facilitate immediate recall, P units are better recalled. Laughery and Pinkus (1968) presented 12 letter strings for .3 sec. to 3.0 sec. and obtained immediate recall. Each string was either M, P, both M and P (words), or randomly ordered letters (R). Results across all presentation rates indicated superior recall for words and P units followed by M units. Recall was poorest for the R strings. Laughery and Pinkus concluded, "pronounceability is a more effective dimension for chunking in STM than meaningfulness [p. 640]." Thus they were suggesting a retention effect due to chunking. Yet in later discussion Laughery and Pinkus attributed the effect to the greater rehearsal efficiency of P units, thereby suggesting an acquisition effect.

The rehearsal efficiency interpretation of the improved recall with P chunks has received considerable support. Stanners and Meunier (1969) found that an unfilled retention interval facilitated immediate recall to the extent that the item string was pronounceable. A similar relation between ease of pronunciation and recall has been obtained by Pinkus and Laughery (1970) and Meunier, Stanners, and Meunier (1971). Gibson, Bishop, Schiff, and Smith (1964) found P units had a lower recognition threshold, which lends further support to a hypothesis that the recall effect is an acquisition phenomenon.

---

[1] This research was supported by Grant APA-373 from the National Research Council to the second author. The report is based on a thesis submitted by the first author in partial fulfillment of the requirements for the master's degree at the University of Manitoba. The authors thank Marilyn Onisko for her assistance in collecting and analyzing the data.

[2] Requests for reprints should be sent to Thomas M. Duffy, Department of Psychology, University of Manitoba, Winnipeg 19, Manitoba, Canada, R3T 2N2.

While the relation between rehearsal efficiency and ease of pronunciation has received considerable support, several studies have found better immediate recall for M rather than P chunks. Boroskin and Lindley (1970) noted that in the Laughery and Pinkus (1968) study P strings always followed the sequence "CVC," while no such consonant-vowel ordering was present in the M strings. With sequential dependency equated, Boroskin and Lindley found greater immediate recall for M strings. Similar results have been obtained by Cimbalo and Mahoney (1970) and Bower and Springston (1970). The basis for the discrepancy in results across studies is unclear. It may be a function of the particular strings employed or a variety of procedural differences (Bower & Springston, 1970). The discrepancy may also be due to a confounding of acquisition and retention. The Gibson et al. (1964) study most clearly tests acquisition. Stanners and Meunier (1969), however, employed an unfilled retention interval and did not equate for level of immediate recall. Thus acquisition and retention were confounded.

In long-term memory (LTM) research, the distinction between acquisition and retention has been adhered to strongly. Underwood (1964) has shown that when level of learning is equated at less than 100%, most variables, like meaningfulness, which had been thought to affect retention were in fact acquisition effects. A similar distinction between acquisition and retention should be made for the STM paradigm. Current models of memory suggest that in the STM paradigm: (*a*) information may be in both STM and LTM and (*b*) information in STM is susceptible to forgetting due to processing of additional input (e.g., Atkinson & Shiffrin, 1968; Waugh & Norman, 1965). Further, information is assumed to transfer to LTM as a function of rehearsal or by the application of coding rules that are based on extraexperimental associations (e.g., remembering a series of three consonants which are your initials). Thus it is possible for two different chunking schemes to be equally effective in immediate recall but to be forgotten at different rates over a filled retention interval due to their differential representation in LTM.

In the present research, level of immediate recall for M, P, and R units was equated by varying the exposure durations of 12 letter strings. The letters were visually grouped into four chunks of three letters, with the chunks being M, P, or R. The exposure durations were determined in an initial experiment in a way analogous to the procedure suggested by Underwood (1964). Thus the rate of acquisition was determined by the exposure duration needed to achieve criterion performance. Retention was measured by the loss over a filled retention interval. For the M and P items, the memory units are likely to be the individual chunks since the letters of each chunk are integrated through their meaningfulness and pronounceability. Since this is well within the capacity of STM, the exposure duration should be relatively short. The research reviewed suggests that P units have a shorter rehearsal cycle than M units, and hence acquisition should occur more quickly for P units. While the information is within the capacity of STM, contact with LTM is made in terms of applying phonemic rules and determining meaning (Bower & Springston, 1970). For the P strings, while the phonemic rules are in LTM, the likelihood of a phonological representation of the three-letter string in LTM is small. Hence, the phonological representation of that string in LTM will be strictly dependent on rehearsal during presentation. This is consistent with the model presented by Laughery (1969). If an M unit is successfully located in the semantic lexicon of LTM, a more permanent contact with LTM is made. That is, "NHL" is already encoded in LTM as "National Hockey League" or as "hockey." It was predicted that a filled retention interval would have a greater effect on the P strings due to their weaker representation in LTM.

Only a superficial coding scheme was provided in the R-string condition. That is, the letters were spatially separated into four groups of three. Thus the letters either had to be rehearsed individually or *S* had to impose an idiosyncratic code. For this reason, it was expected that exposure duration would be considerably longer for these strings than for the M and P strings in order to achieve criterion performance. Relatively little retention loss, however, should occur for the R strings. If rehearsal is by individual letters, then a greater proportion of those letters (relative to the chunked M and P strings) must be transferred to LTM to obtain an equivalent level of immediate recall. This is due to the limited STM capacity. A filled retention interval should have no effect on the LTM items (Glanzer & Cunitz, 1966), and recall should therefore be higher for M and R strings than for P strings. If idiosyncratic encoding occurs, it must be semantically based (the letters are not pronounceable), and this information is already in LTM. Thus, again recall is from both STM and LTM, and delayed recall should be greater relative to P strings.

METHOD

Experimental Design

Experiment I.—This experiment was designed to determine the acquisition rate of M, P, and R strings. The string-type variable was manipulated between *Ss*. Exposure duration was a within-*S* factor, and variations in it were determined by each *S*'s performance. That is, the duration was increased or decreased by a fixed amount on Trial N, depending on whether *S* recalled fewer or more than a criterion of six to nine letters on Trial N−1. Each *S* received nine trials with a different string presented on each trial. The order of presentation of strings was randomly determined for each *S*. The criterion exposure duration for *S* was the mean duration on those trials on which criterion recall was attained. Presentation time for Exp. II was then determined by averaging the criterion exposure durations across *Ss*.

Experiment II.—This experiment involved a 3 × 3 × 2 mixed-model factorial design. The major experimental variables were string type (M, P, or R) and retention interval (0, 5, or 10 sec.). String type was manipulated between *Ss*, while retention interval was a within-*S* factor. The third factor involved the use of two *Es*. Each *S* received nine trials, three at each retention interval. Each string was presented with each retention interval to an equal number of *Ss*. There were two random orders of presentation of retention interval conditions.

Materials

Each string type consisted of 12 letters grouped into units of 3. Nine strings of each type were constructed with one restriction being that no letter occurred more than three times in a string. The P strings were taken from Boroskin and Lindley (1970), who scaled the units for degree of pronounceability. There was no consistent ordering of consonants and vowels, e.g., BLI VEA SLE ORC. The M strings each consisted of four 3-letter acronyms found to be generally meaningful to students at the University of Manitoba. In a departure from previous research, only consonant strings were used, e.g., NFL CBS LTD MPH. It was felt that the addition of any vowels would result in some pronunciation and could indeed lead to idiosyncratic regrouping of the letters into pronounceable units. The R strings also were constructed without vowels and consisted simply of a random reordering of the letters in the M strings.

The stimulus letters were placed on white cards using Letterset type $\frac{3}{4}$ in. high. There was a $\frac{1}{4}$-in. space between each group of three letters and $\frac{1}{8}$ in. between letters within a group. The strings were presented on a three-channel Scientific Prototype tachistoscope.

Subjects

One hundred fifty-two students enrolled in summer session psychology courses served as *Ss*. Participation was voluntary, but approximately 40% received some course credit. These *Ss* were evenly distributed over experimental conditions. Twenty-four *Ss* took part in Exp. I (8 per condition) and 128 in Exp. II (14 per condition).

Procedure

Experiment I.—The *Ss* were run individually and were assigned to string-type conditions following a predetermined random order. The first trial for each condition involved a presentation time which preliminary work indicated would yield somewhat less than criterion performance. These times were 3.5 sec., 3.5 sec., and 15 sec. for Cond. P, M, and R, respectively. Criterion performance on a trial was six to nine letters correct. On each trial, the presentation time was adjusted by a fixed interval depending on *S*'s performance on the prior trial. The interval of adjustment was .2 sec. for Cond. P and M and 2.0 sec. for Cond. R.

The *Ss* were instructed as to the nature of the letter strings. That is, they were told that a sequence involved 12 letters grouped into four 3-letter units that were meaningful, pronounceable, or trigrams (for the respective conditions). They were also told that remembering the units would be easier than trying to remember the letters individually. Recall was written and it was explained that any letter could be written down first but that the sequence on the sheet must be properly ordered. The onset of each trial was signaled by a verbal

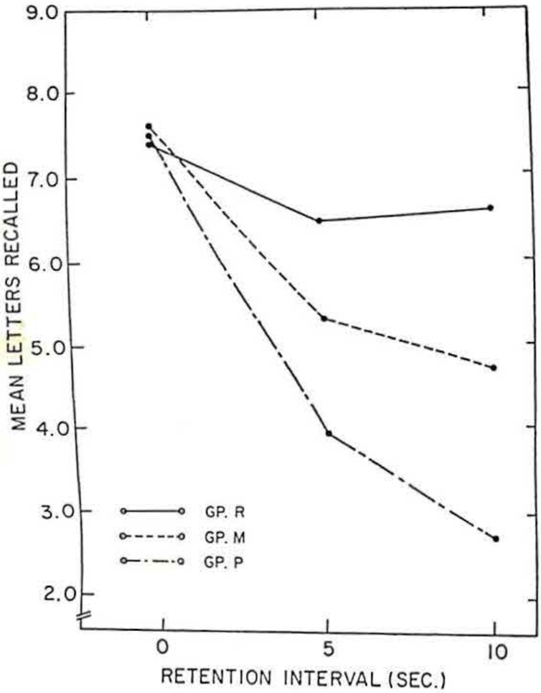


FIG. 1. Mean letters recalled as a function of retention interval and string type (Exp. II).

"Ready" 3 sec. prior to presentation, and recall was signaled by the end of presentation and by handing *S* a recall sheet. Recall time was 30 sec.

Experiment II.—The *S*s were randomly assigned to the six conditions obtained by using all combinations of string type (M, P, and R) and *E*s (1 and 2). Instructions and mode of presentation were the same as in Exp. I except that the retention interval activity was explained. The onset of each of the nine trials was signaled by a verbal "Ready" 3 sec. prior to presentation. The string was then presented for a time as determined in Exp. I. The end of the presentation was followed by either a verbal "Recall" or a Stroop test card (Underwood, 1967, Ch. 12). The Stroop card was placed in front of *S*, and he was required to read the color names as quickly as possible for the duration of the retention interval. There was a 35-sec. recall period for written recall, and this was followed by a 20-sec. intertrial interval.

Following the nine STM trials, each string was presented on a sheet of paper and *S*s were asked to provide a protocol indicating any techniques used in remembering items. For the M condition, *S*s were also asked if they knew the meaning of each acronym.

RESULTS

Experiment I.—The average presentation time on those trials resulting in criterion performance (six to nine letters correct) was calculated for each *S*. The means of these times were 3.1 sec., 3.8 sec., and 17.5 sec. for Cond. P, M, and R, respectively. A two-tailed *t* test revealed a highly significant difference between the P and M groups, $t\ (26) = 5.29,\ p < .002$. Of practical significance, the longest criterion presentation time for an *S* in the P condition was shorter than the time for every *S* in the M group. Clearly the difference between the M and P groups and the R group is highly significant.

Since criterion performance involved six to nine letters correct, it is possible that the shorter presentation time for Group P might have been correlated with fewer letters correct. However, the results indicate that this was not the case. On those trials involving criterion recall, the mean letters recalled were 7.3, 7.4, and 7.3 for Cond. P, M, and R, respectively. An analysis of variance (ANOVA) of these scores indicated the differences were not significant ($F < 1$).

Experiment II.—The mean number of letters recalled in proper serial position as a function of string type and retention interval is presented in Fig. 1. Looking only at the immediate recall scores, it is clear that level of learning was successfully equated. The obtained means were 7.5, 7.6, and 7.4 for Cond. P, M, and R, respectively. An ANOVA performed on these scores failed to yield a significant difference ($F < 1$).

A 3 × 3 × 2 mixed-model ANOVA was performed on all of the recall scores. The results indicated significant main effects due to string type, $F\ (2, 36) = 5.62,\ p <$ .01, and retention interval, $F\ (2, 72) = 5.3,$ $p < .001$. The only significant interactions were String Type × Retention Interval, $F\ (4, 72) = 7.16,\ p < .001$, and E × Retention Interval, $F\ (2, 72) = 3.43,\ p < .05$. The interaction involving *E*s is due to a greater retention loss obtained by one of the *E*s. As can be seen in Fig. 1, the other effects are due to the differential retention loss as a function of string type. Retention loss was greatest for the P strings and least for the R strings. A post hoc comparison of string-type conditions at the 10-sec. retention interval was made using the Tukey HSD test. The results indicated that Group M recalled significantly fewer letters than Group R ($p < .01$) and significantly more letters than Group P ($p < .05$).

At the end of the experimental session *S*s in Group M were asked the meaning of each acronym. Only 13.6% (69/504) of the acronyms were reported as unknown. The effect of meaning on retention can be seen clearly in the retention of these acronyms relative to those where meaning was known. When the acronyms were reported as known, the proportion of letters recalled at the 0- and 10-sec. retention intervals was .657 and .420, respectively, while the proportions are .509 and .108 for acronyms with meaning reported as unknown. Thus the loss over the retention interval was least for those acronyms for which information as to meaning was represented in LTM. Thus knowing the meaning of the acronym facilitated immediate recall and reduced retention loss.

DISCUSSION

The acquisition results clearly support Laughery and Pinkus (1968) in that presentation time was shorter, and hence acquisition was more rapid for pronounceable units. The basis of the discrepancy between these results and those obtained by Boroskin and Lindley (1970), in which immediate recall was greater for meaningful units, is unclear. However, it should be noted that pronounceability was found to be the more effective acquisition unit in both the present data and the Laughery and Pinkus (1968) findings, even though the studies differed in terms of acquisition measures, presentation methods, and restrictions in item construction.

The major finding of the present research is a complete reversal of relationship between string types when one goes from acquisition to retention. In terms of delayed recall, the meaningful strings were recalled better than the pronounceable strings. However, it is questionable whether one would conclude that meaningfulness is a better means of encoding. First, the statement would have to be restricted to delayed recall. Second, random strings, for which no coding scheme was provided, showed the least retention loss.

An interpretation based on the contributions of STM and LTM seems appropriate. The present data suggest that the more superficial the code (i.e., the less it depends on semantics) the less contact is made with LTM. Successful recall is therefore dependent largely on STM for superficially coded items. Thus a filled retention interval should produce a greater loss for these items. If, however, additional rehearsal time is allowed, as was the case for R strings, then the superficial code should transfer to LTM. Thus retention loss is reduced by either semantic coding or rehearsal time of sufficient length to allow transfer to LTM.

Acquisition or immediate recall should be unaffected by the extent of semantic content in the code. Rather acquisition should only be dependent on the time needed to identify or process the code. Consistent with this, Gibson et al. (1964) found a lower identification threshold for pronounceable units. It should be possible, however, to find meaningful items which are more easily recognizable than the items employed due to their recency of occurrence in extraexperimental situations. This would result in a lower recognition threshold for M strings as compared to P strings. The relationship in delayed recall should be unaffected, however, since the nature of the code with respect to LTM representation is unchanged.

That a nonsemantically based code is a temporary and superficial code is consistent with Laughery (1969). He suggests that whereas the meaningful unit is already established in LTM, the pronounceable unit involves an unfamiliar combination of phonological rules and thus must be established in LTM via rehearsal. If sufficient rehearsal is not allowed, then the pronounceable unit should be more susceptible to forgetting when a rehearsal-preventing task is introduced. The data of the present study confirm this notion. Interference based on specific similarity relations, however, should affect both pronounceable and meaningful units equally. Indeed, in the present research many *S*s recalled "NBC" as "CBS," suggesting a confusion based on the semantic tag "television station." While acoustic errors could not be assessed it is likely that such errors occurred in the recall of the pronounceable strings.

A final implication of the present findings relates to Underwood's (1964) finding that equating level of learning nullifies the effect of meaningfulness in retention. This has been shown to be the case in the LTM paradigm.

In the present study, differential forgetting based on meaningfulness was obtained in spite of the care taken to equate level of learning. It would seem that in the STM paradigm, equating for level of immediate recall results in strings having differing proportions of items in STM and LTM. The differential rate of forgetting can be accounted for by the unequal representation in the memory systems. Thus the present data support a multimemory system. In addition, the results indicate the dilemma faced in attempting to examine forgetting in STM. That is, failure to equate for level of learning results in a confounding of learning and forgetting. Equating for level of learning, however, results in a confounding of forgetting and representation in STM.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 3. New York: Academic Press, 1968.

Boroskin, A., & Lindley, R. H. Meaningfulness versus pronunciability in immediate memory and free recall. *Journal of Experimental Psychology*, 1970, **83**, 182–184.

Bower, G. H., & Springston, F. Pauses as recording points in letter series. *Journal of Experimental Psychology*, 1970, **83**, 421–430.

Bower, G. H., & Winzenz, D. Comparison of associative learning strategies. *Psychonomic Science*, 1970, **20**, 118–119.

Cimbalo, R. S., & Mahoney, M. S. Short-term memory: Effects of type of trigram isolated and position of isolation. *Psychonomic Science*, 1970, **18**, 77–78.

Gibson, E. J., Bishop, C. H., Schiff, W., & Smith, J. Comparison of meaningfulness and pronunciability as grouping principles in the perception and retention of verbal material. *Journal of Experimental Psychology*, 1964, **67**, 173–182.

Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

Laughery, K. R. Computer simulation of short-term memory: A component-decay model. In G. H. Bower & J. T. Spence (Eds.), *The Psychology of learning and motivation: Advances in research and theory.* Vol. 3. New York: Academic Press, 1969.

Laughery, K. R., & Pinkus, A. L. Recording and presentation rate in short-term memory. *Journal of Experimental Psychology*, 1968, **76**, 636–641.

Meunier, G. F., Stanners, R. F., & Meunier, J. A. Pronounceability, rehearsal time, and the primary effect of free recall. *Journal of Experimental Psychology*, 1971, **88**, 123–127.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

Pinkus, A. L., & Laughery, K. R. Recoding and grouping processes in short-term memory: Effects of subject-paced presentation. *Journal of Experimental Psychology*, 1970, **83**, 335–341.

Stanners, R. F., & Meunier, G. F. Pronounceability and rehearsal time in STM with controlled acquisition. *Journal of Experimental Psychology*, 1969, **80**, 359–363.

Underwood, B. J. Degree of learning and measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 112–129.

Underwood, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1967.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

(Received February 28, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 297–300

ROLE OF STIMULUS COMPARISON IN EQUIVALENCE TRAINING[1]

DAVID R. THOMAS,[2] JAMES T. MILLER, AND GARY HANSEN
University of Colorado

Four groups of 10 pigeons each pecked a key for variable internal (VI) reinforcement. The Single Stimulus group saw only the green training stimulus. The One Switch group received 7 days of training with green followed by 7 days of of training with a white vertical line. For the One Switch per Day group, green was present during the first half of each daily session, with line present during the last. In the Random Alternation group, the green and line stimuli were alternated randomly in 1-min. periods. All groups then received a wavelength generalization test in extinction. The Single Stimulus group and One Switch group gradients were quite similar and were sharper than the One Switch per Day group and Random Alternation group gradients, which also were alike. Thus, in nondifferential training some opportunity for comparison of the training stimuli is necessary to flatten the resulting generalization gradient, although that opportunity need not be extensive.

Discrimination training and nondifferential ("equivalence") training may be viewed as complementary procedures to the extent that they influence stimulus control in opposite directions. Discrimination training sharpens generalization gradients; nondifferential training, on the other hand, flattens them (cf. Lyons & Thomas, 1967; Switalski, Lyons, & Thomas, 1966).

In a recent study, Ellis (1970) investigated the effects on stimulus generalization of both type and number of stimulus sequences administered during intradimensional discrimination training. One of Ellis' groups received nine sessions of training with an S+ followed by a single S− session and a subsequent generalization test in extinction. This one switch between S+ and S− conditions was not sufficient to produce the typical effects of intradimensional discrimination training, i.e., sharpening of the gradient, behavioral contrast, and the peak shift. This group replicated an earlier finding by Honig, Thomas, and Guttman (1959). Ellis found, however, that *S*s who received one S−/S+ comparison on each of 10 days (i.e., with a block of S− periods preceding a block of S+ periods) yielded generalization gradients comparable to those of *S*s for which S+ and S− stimuli were alternated randomly. Clearly the effects of discrimination training on stimulus generalization require more than the reinforcement of responding in the presence of one stimulus and extinction in the presence of another. Some comparison or alternation of the two stimuli is required, although the number of opportunities for comparison may be quite limited. This finding is inconsistent with simple S–R associationistic views of discrimination learning (cf. Spence, 1936), which ascribe performance to the interaction of hypothetical excitatory and inhibitory response tendencies. It favors a more cognitive view of the learning process, such as that espoused by Thomas (1969, 1970). Thomas has proposed that discrimination training establishes a state of "general attentiveness," such that training in which one stimulus difference is significant enhances attentiveness to a wide range of stimuli. By the same token, nondifferential training is thought to teach *S* that stimuli are insignificant, reducing attention to stimulus change and thereby flattening gradients of stimulus generalization. If a common attentional mechanism underlies the effects on generalization of both discrimination and nondifferential training, then opportunity to compare the training stimuli should play a parallel role in these two paradigms. The present experiment was designed to test this hypothesis.[3]

---

[1] This study was supported by National Institutes of Health Research Grant HD-03486 and Training Grant MH-10427 under the direction of David R. Thomas.

[2] Requests for reprints should be sent to David R. Thomas, Department of Psychology, University of Colorado, Boulder, Colorado 80302.

Method

Subjects.—Forty domestic, naive pigeons obtained from a local supplier and maintained at 70%–75% of free-feeding weight were used in the experiment.

Apparatus.—Each *S* was run in one of five identical operant conditioning chambers enclosed within individual sound-attenuating hulls. Stimuli were projected onto the response key by IEE in-line projectors Model 10-3043-1851-L, with General Electric 1813 bulbs. Five chromatic stimuli, with peak wavelengths at 501, 538, 555, 576, and 606 nm., and a vertical line .32 × 2.22 cm. on a black background, were available for projection onto the response key. All light within the chamber was provided by the response key, except during 3-sec. reinforcement periods when the food hopper was illuminated. Extraneous noise was masked by "white noise" introduced into the chamber by a small speaker located on the wall near the response key.

Procedure.—Prior to training, all *S*s were assigned randomly to one of four groups, which were designated Single Stimulus, One Switch, One Switch per Day, and Random Alternation. Initial training for all *S*s was identical, with the 555-nm. stimulus present on the respense key. Both magazine training and key-peck training were accomplished on the first day. During the second and third of the daily training sessions, 30 continuously reinforced key pecks were permitted. During the fourth, fifth, and sixth session, gradually increasing variable interval (VI) requirements were programmed such that by the end of the sixth session, a variable interval 1-min. (VI-1 min.) schedule was in effect. On the next day, the experimental condition for each group was introduced.

Fourteen training sessions with the experimental conditions were run. Each of these sessions was 30 min. in duration with a VI-2 min. schedule of reinforcement in effect, regardless of the stimulus projected onto the key. The *S*s in the Single Stimulus group simply were given additional training with the 555-nm. stimulus. The *S*s in the One Switch group had the 555-nm. stimulus projected onto the key during the first seven of the training sessions and the white vertical line stimulus during the last seven. Those *S*s in the One Switch per Day group experienced the 555-nm. stimulus for the first 15 min. of each session and the white vertical line during the last 15 min. The Random Alternation group received 1-min. presentations of these same two stimuli, which were ordered randomly with the restriction that neither stimulus appeared more than three times in succession. No blackouts intervened between successive stimulus presentations for any group.

On the day following the 14 VI-2 min. training sessions, all *S*s were given a stimulus generalization test in extinction. The test consisted of 12 blocks of five chromatic stimuli, 501, 538, 555, 576, and 606 nm., randomly presented within each block. Each stimulus was presented for 1 min. with no blackouts between successive stimulus presentations.

Results

For reasons which need not be repeated here, relative generalization gradients have become the datum of choice in operant generalization studies (cf. Hearst, 1969; Honig, 1969; Thomas & King, 1959). The total number of responses made to each of the test stimuli was divided by the total

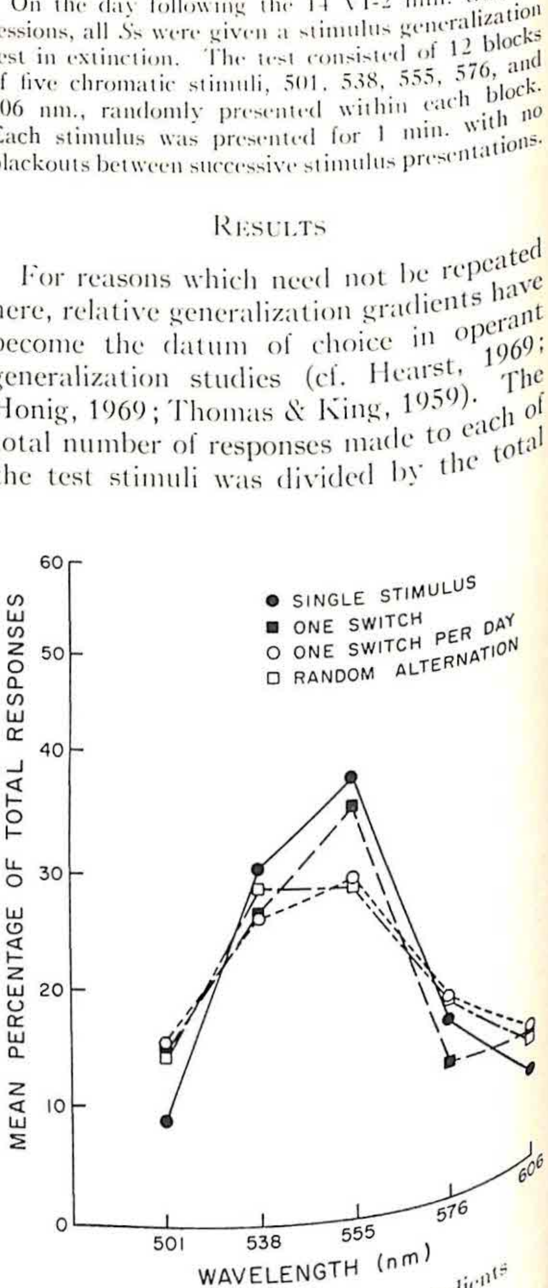


FIG. 1. Relative generalization gradients of the four groups of *S*s.

---

[3] It should be noted that the concept of "general attentiveness" does not provide a complete account of performance following intradimensional discrimination training, since it provides no basis for predicting the peak shift or behavioral contrast. Thus for present purposes, the only finding of the Ellis (1970) experiment which is relevant is that concerning differences in the slope of postdiscrimination gradients.

number of responses during the entire test to obtain a gradient of relative generalization for each *S*. The mean relative gradient for each group is shown in Fig. 1. Inspection of the figure suggests a confirmation of the experimental hypothesis. The gradients of the Single Stimulus and One Switch groups are very similar and they appear sharper than the gradients of the One Switch per Day and Random Alternation groups, which virtually are superimposed. For statistical purposes, for each *S* the percentage of total responses emitted to the 555-nm. training stimulus was used as an index of generalization slope. Within each of the two pairs of groups, *t* tests were performed on these slope scores, and in neither case did the *t* approach significance ($t < 1$). This permitted a pooling of *S*s within each pair, and a *t* test between Single Stimulus and One Switch groups (combined) and One Switch per Day and Random Alternation groups (combined) was performed: $t\ (37) = 2.64$, $p < .02$.[4]

In the present study, just one switch between the two training stimuli was insufficient to reliably flatten the gradient, relative to the Single Stimulus control. In addition to experience with two reinforced stimuli, rather than one, there are some other reasons why the One Switch group *might* have produced a flatter gradient. The One Switch group was tested 8 days after last exposure to the 555-nm. training stimulus, whereas the Single Stimulus group was tested after a 1-day delay. This gave the One Switch group much more opportunity to "forget" the value of the training stimulus. However, Thomas and Lopez (1962) found that generalization gradients flatten with the introduction of a 1-day delay between training and testing, but they found little or no additional flattening with a 1-wk. delay.

A more important procedural difference between Single Stimulus and One Switch groups is the fact that the latter had 7 days of VI-2 min. training with the 555-nm. stimulus rather than 14. Hearst and Koresko (1968) have shown that relative generalization gradients sharpen with increased training. *If* the One Switch gradient had been reliably flatter than the Single Stimulus control, the difference in amount of training might have accounted for it. The absence of a significant difference, however, strongly argues for the role of stimulus comparison in producing nondifferential training effects.

Using the same interdimensional task employed here, Switalski et al. (1966) and Lyons and Thomas (1967) have demonstrated the flattening of generalization by nondifferential training. These prior studies both employed a within-*S*s design. As far as we know, the present study is the first between-group experiment to demonstrate a flattening of generalization by interdimensional nondifferential training. This attests to the power and the generality of nondifferential training effects.

Ellis (1970) had reported that one S−/S+ switch per day was sufficient to produce the typical effects of discrimination training. In a parallel fashion, the present study demonstrates that one switch per day produces about as much flattening of the gradient as does random alternation. Thus stimulus comparison seems to play a comparable role in discrimination and nondifferential training. This evidence supports the notion that a single underlying mechanism, such as that of "general attentiveness," determines the effects of these two training procedures on the slope of generalization gradients.

REFERENCES

ELLIS, W. R., III Role of stimulus sequences in stimulus discrimination and stimulus generalization. *Journal of Experimental Psychology*, 1970, **83**, 155–163.

HEARST, E. Excitation, inhibition, and discrimination learning. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental issues in associative learning*. Halifax, Nova Scotia: Dalhousie University Press, 1969.

HEARST, E., & KORESKO, M. B. Stimulus generalization and amount of prior training on variable interval reinforcement. *Journal of Comparative and Physiological Psychology*, 1968, **66**, 133–158.

HONIG, W. K. Attentional factors governing the slope of the generalization gradient. In R. M.

---

[4] One *S* in the One Switch per Day group had to be discarded because of apparatus failure during testing, reducing the *n* to 9 in that group, and the *df* to 37 in the *t* test.

Gilbert & N. S. Sutherland (Eds.), *Animal discrimination learning*. London: Academic Press, 1969.

HONIG, W. K., THOMAS, D. R., & GUTTMAN, N. Differential effects of continuous extinction and discrimination training on the generalization gradient. *Journal of Experimental Psychology*, 1959, **58**, 145–152.

LYONS, J., & THOMAS, D. R. Effects of interdimensional training on stimulus generalization: II. Within-subjects design. *Journal of Experimental Psychology*, 1967, **75**, 572–574.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychological Review*, 1936, **43**, 427–449.

SWITALSKI, R. W., LYONS, J., & THOMAS, D. R. Effects of interdimensional training on stimulus generalization. *Journal of Experimental Psychology*, 1966, **72**, 661–666.

THOMAS, D. R. The use of operant conditioning procedures to investigate perceptual processes in animals. In R. M. Gilbert & N. S. Sutherland (Eds.), *Animal discrimination learning*. London: Academic Press, 1969.

THOMAS, D. R. Stimulus selection, attention, and related matters. In J. H. Reynierse (Ed.), *Current issues in animal learning*. Lincoln: University of Nebraska Press, 1970.

THOMAS, D. R., & KING, R. A. Stimulus generalization as a function of the level of motivation. *Journal of Experimental Psychology*, 1959, **57**, 323–328.

THOMAS, D. R., & LOPEZ, L. J. The effects of delayed testing on generalization slope. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 541–544.

(Received February 28, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 301–306

SENSE IMPRESSION AS AN ENCODING DIMENSION OF WORDS[1]

DELOS D. WICKENS[2]
Ohio State University

DONALD B. REUTENER
Smith College

F. THOMAS EGGEMEIER
United States Army Natick Laboratories, Natick, Massachusetts

Five experiments were conducted to measure release from proactive inhibition in the L. R. Peterson and M. J. Peterson paradigm as a function of changes in various subdimensions of the broad dimension of sense impression. It was found that certain aspects of this general dimension seemed to serve as quite effective encoding attributes, while others were less potent. An extradimensional shift, such as round and white, was quite powerful, whereas the intradimensional change of round and rectangular was the weakest of the group. Other intradimensional shifts (large and small; hard or soft) were of intermediate value. It is concluded that the attribute of sense impression can be a significant encoding dimension of words.

Underwood and Richardson (1956), in the process of seeking sets of words to which *S*s would give a common response, discovered that their purpose was attained when *S*s gave responses which *E*s described as "sense impressions." These *E*s then went on and specifically directed and guided a new group of *S*s to give sense-impression responses to a large number of words, and these words and the sense impression they provoked—such as round, small, white, etc.—constitute the major bulk of the article.

Since that time, several experiments have used these sense-impression words as an independent variable with varying degrees of success. Underwood and Richardson (1956b) employed them as materials for a concept-learning study and found that their use facilitated concept learning, particularly for words which were of high dominance level, and particularly when *S*s were rather completely instructed on their potential usefulness. On the other hand, Richardson (1962) in a transfer type of situation found essentially no effect of this variable. Two studies (Bousfield & Puff, 1964; Hudson, 1968) show that, in the free-recall situation, clustering based upon sense impression does occur.

Recent theoretical discussions (Bower, 1967; Morton, 1970; Norman & Rumelhart, 1970; Underwood, 1969; Wickens, 1970) on the nature of word meaning have suggested that total word meaning is comprised of a number of different attributes, and an obvious task for the experimentalist is to identify these attributes and determine their relative dominance in the encoding process. The manner in which the sense-impression attribute was discovered and the outcome of several of the above-mentioned experiments indicates that sense impression is an attribute of word encoding. This fact seems to justify the desirability of exploring the sense-impression attribute by another method which can give a quantitative estimate of an attribute's effectiveness and which will permit it to be compared with other attributes in saliency. The method in question is the "release from proactive inhibition (PI)" technique which will be briefly described below.

The method is presented in more detail elsewhere (Wickens, 1970), but a brief description of it is in order at this time. Using the Peterson and Peterson (1959) paradigm, triads of words drawn from the same class are presented for three trials at a 20-sec. retention interval. On the fourth trial, the triad for the experimental group is drawn from a different class. The control group receives the same class of material for all four trials. The typical result during the first three trials is a decline in performance from about 90% correct on Trial 1 to 40% on Trial 3. The Control group will drop a few percentage points more on Trial 4; but, depending upon the class of items being investigated, the performance of the Experimental group will increase to a greater or lesser degree. When increments occur, they are assumed to occur because the materials of the first three trials are encoded differently than those of the fourth trial. This assumption leads to the interpretation that one has identified an attribute of psychological significance for *S* (Wickens, 1970). The present article describes the results derived from applying this technique to various sense impression dimensions.

---

[1] This research was supported in part by Grant OEG-5-9-450276 from the United States Office of Education.

[2] Requests for reprints should be sent to Delos D. Wickens, Department of Psychology, 1945 North High Street, Ohio State University, Columbus, Ohio 43210.

EXPERIMENT I

Method

Subjects.—The *S*s in this experiment were 96 Ohio State University introductory psychology students who chose this experiment to meet part of the course requirement. About two-thirds of them were females and the proportion of the two sexes was approximately equal in the various groups of the experiment.

Materials.—The materials employed in the experiment were taken from the high-dominance words of the round and white sense impressions of Underwood and Richardson (1956a). Specifically the triads were as follows: *round*—barrel, knob, doughnut; head, balloon, wheel; globe, baseball, spool; dome, saucer, button and *white*—milk, teeth, napkin; linen, rice, bandage; chalk, lint, salt; snow, collar, diaper.

Procedure.—After *S* came into the experimental room, *E* told *S* that he was interested in *S*'s ability to remember three words for a short time and also in his competence in counting backwards by three's. He was further told that both of these tasks were of equal importance and that he should do as well in each of them as he could.

The *S* was then given an example of the timing and sequence of the experiment. This sequence consisted of a warning asterisk for 2 sec., a 2-sec. exposure of the triad (in this sample trial it was a blank slide), a two-digit number from which *S* counted backwards by 3's to the 1-sec. beat of a metronome, a question mark which was the signal to recall 20 sec. after the presentation slide went off, an 8-sec. recall period, and the asterisk warning for the new trial. If *S* made it clear that the procedure was understood, the experiment was begun.

There were two major groups in the experiment, the Control and the Experimental. Each of these consisted of two subgroups. Half of the Controls experienced round-word triads throughout the four trials, and the other half white-word triads. Half of the Experimental group received round-word triads for the first three trials and then was shifted to white-word triads. The reverse arrangement applied to the other Experimental subgroup. The particular slides were rotated across position so that each slide appeared equally often in each trial position. Except for equating for sex, *S*s were assigned to groups in order of appearance in the laboratory.

Results

The scoring system used was to give one point for each item correctly recalled and one additional point if all three were recalled in the correct order. Scores, therefore, could range from 4 to 0. Parenthetically, it should be added, that in several of these types of experiments we have compared this method with the method of giving points only for items recalled. The conclusions that could be drawn from the data have always been the same at about

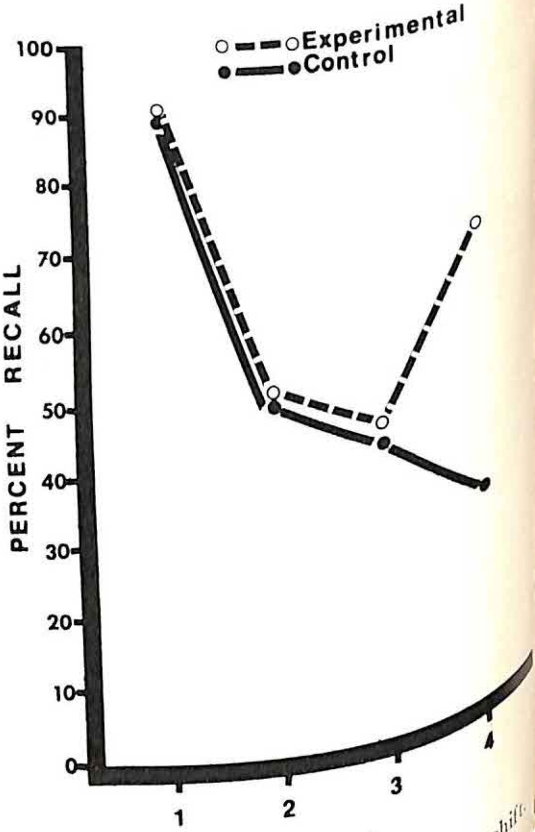


FIG. 1. Performance on the round–white shift

the same statistical level. The present method gives a little, but not much, credit for order information.

The results of the experiment for the combined major groups of the experiment are presented in Fig. 1. An analysis of variance was conducted across the first three trials to test for the buildup of proactive inhibition and to determine whether the performances of the Experimental and Control groups were essentially alike. The trials effect was highly significant, F (2, 184) = 100.9, $p < .001$, and the F for the Experimental–Control comparison was less than 1.0. A difference which is not relevant to the major purpose of the experiment was found in the two classes of materials with the white-word triads being somewhat easier than the round-word triads, F (1, 92) = 10.16, $p < .001$. It will be recalled that classes are counterbalanced across groups. Finally, a planned t test was run between the Experimental and Control groups on Trial 4 and the result obtained was t (95) = 5.54, $p < .001$.

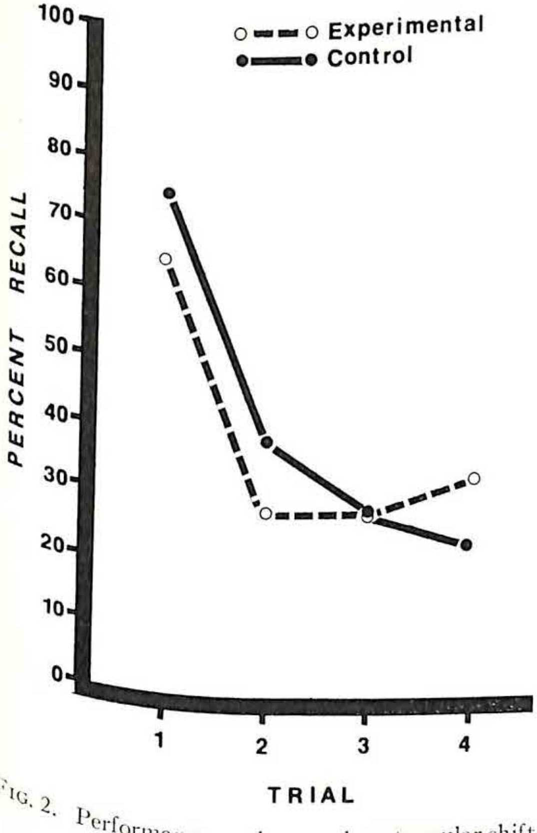


FIG. 2. Performance on the round–rectangular shift.

EXPERIMENTS II, III, IV, AND V

Method

Subjects.—There were 96 *S*s in each of these experiments, drawn from the same population and assigned to groups in the same manner as were those of Exp. I.

Materials.—It requires at least 12 words of a given class to conduct an experiment of the sort described above; and, although this number was available for the round and white attributes in the Underwood and Richardson (1956) report, their study did not seem to include enough items to test other sense-impression dimensions. For this reason, a group of approximately 75 *S*s were asked to list as many exemplars as they could of the following attributes: square–rectangular; hard, soft, large, small, heavy, and light. These data were then tabulated and the 12 words with the highest frequency range were chosen to be used as materials in the subsequent experiments. It will be noted that this method of obtaining examples is like that used by Battig and Montague (1969) with taxonomic categories in that the category was given and a request was made for examples. Underwood and Richardson worked the other way around, giving words and asking for the sense impression that the words created.

The words generated gave the materials to conduct encoding tests along the following dimensions: rectangular–round, hard–soft, large–small, and heavy–light.

The particular triads used in each case were *rectangular*—box, poster, table; envelope, door, brick; book, dice, photograph; window, dollar, bed; *round*—the same as in Exp. I; *hard*—rock, fist, enamel; steel, bone, marble; diamond, brick, iron; cement, ice, hammer; *soft*—feather, lips, cotton; cushion, pudding, girl; pillow, fur, baby; kitten, bed, velvet; *large*—ocean, truck, sky; elephant, sun, giant; mountain, world, ship; universe, whale, building; *small*—mouse, pin, flea; baby, molecule, cell; germ, needle, ant; atom, dust, splinter; *heavy*—boulder, whale, steel; anchor, car, elephant; ton, building, airplane; wrestler, gold, truck; *light*—feather, leaf, cotton; cloud, pin, ashes; ounce, aluminum, thread; dust, bubble, snow.

The attributes tested in each experiment were: for Exp. II, rectangular and round; for Exp. III, hard and soft; for Exp. IV, large and small; and for Exp. V, heavy and light.

Procedure.—The procedure of these experiments was the same as that in Exp. I. In addition, the method of scoring the data was the same.

Results

The results for Exp. II, the rectangular–round shift, are presented in Fig. 2. Statistical analysis for the first three trials showed only a significant trials effect, F (2, 184) = 82.3, $p < .001$. The t value

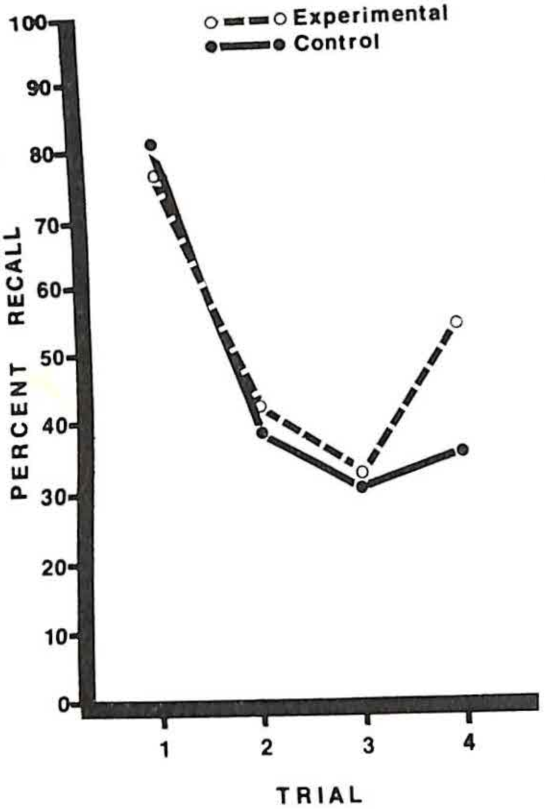


FIG. 3. Performance on the hard–soft shift.

was t (95) = 1.76, $p < .10$, for the fourth trial.

The results for Exp. III, the hard–soft shift, are presented in Fig. 3. Again only the trials effect was significant across the first three trials, $F(2, 184) = 95.9$, $p < .001$. In this instance, the fourth trial t was t (95) = 3.32, $p < .001$.

The results for Exp. IV, the large–small shift, are presented in Fig. 4. The trials effect across the first three trials was F (2, 184) = 69.9, $p < .001$, and nothing else was significant. In addition, the fourth trial t was highly significant, t (95) = 4.18, $p < .001$.

The results for Exp. V, the heavy–light shift, are presented in Fig. 5. Across the first three trials, the analysis of variance produced an F (2, 184) = 52.5, $p < .001$, for the trials effect with nothing else significant. On Trial 4 the t was nonsignificant at 1.14.

DISCUSSION

A way of quantifying the results making it possible to compare one type of shift with another is to obtain what has been called the percentage release from proactive inhibition (PI). This is calculated by measuring the total decline for the Control group from Trial 1 through Trial 4 and determining the amount by which the Experimental group excels the Control on Trial 4. The percentage that this latter value is of the total control decline can be called the percentage release from PI.

When these percentages are calculated for the present five experiments, they give the following accounting: 67% for round–white, 20% for rectangular–round, 42% for hard–soft, 50% for large–small, and 17% for heavy–light.

It is quite apparent from these data that the sense impression created by a word can serve as an attribute of a word in the automatic and spontaneous type of encoding which seems to prevail in this type of situation. It is also obvious that its effectiveness varies across attributes of the sense-impression dimension.

The greatest release was found in the round–white shift, and it should be noted that this

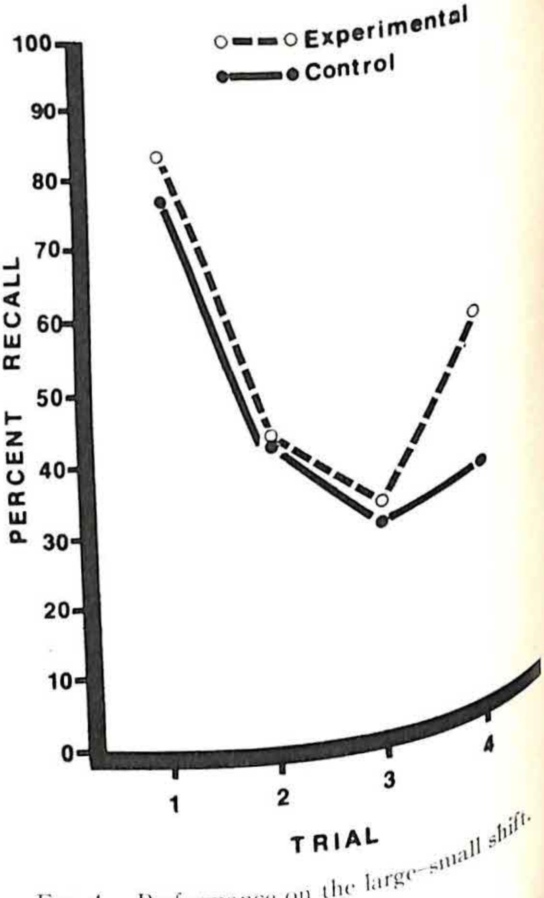


FIG. 4. Performance on the large–small shift.

may be described as an extradimensional shift, namely from color to shape. When one shifts within the form dimension alone, as in Exp. II, the effect is slight. That intra- versus extradimensional shifts are not, however, the critical conditions for release from PI is indicated by the sizable amounts obtained from going from one to the other end of the hard–soft and the large–small continuum. Why there should be a variation from 17% to 50% across the intradimensional shift is unclear. Indeed, some of the same words occur in the most effective dimension—large and small—and in the least effective dimension—heavy–light. One can only infer from the data that, for whatever reason, some dimensions are more salient in their encoding characteristics than are others.

The amount of release which can be obtained by the manipulation of at least some of the sense-impression dimensions is quite sizable as compared with other types of shift. In a study by Goggin and Wickens (1971), Spanish–English shifts with bilinguals were executed leading to a percentage release of about 70%; and in the same study a taxonomic category shift (body parts and food names) resulted in an increment of a slightly larger value. Thus the extradimensional shift in Exp. I is quite impressive as are the intradimensional shifts of hard–soft and large–small. The rather high degree of effectiveness of the sense-impression dimension in the present experiments stands in contrast to the ineffectiveness of it in the transfer situation of Richardson (1962). In the Richardson experiment, taxonomic category proved to be a potent source of transfer, but sense impression did not do so. The data of the present experiment find sense impressions to be strong encoding dimensions, though generally less potent than taxonomic category (Goggin & Wickens, 1971; Loess, 1967). Similarly, sense impression does not prove to be very important as a carrier of concept information in the Underwood and Richardson (1956b) experiment. The reason for the variation in effectiveness across these experiments is not clear, but it may be that *S* adopts a strategy in the concept and transfer experiments which obscures the cues available in a more "natural" state of behavior. Certainly sense impression is effective as an encoding dimension in the release from PI situation.

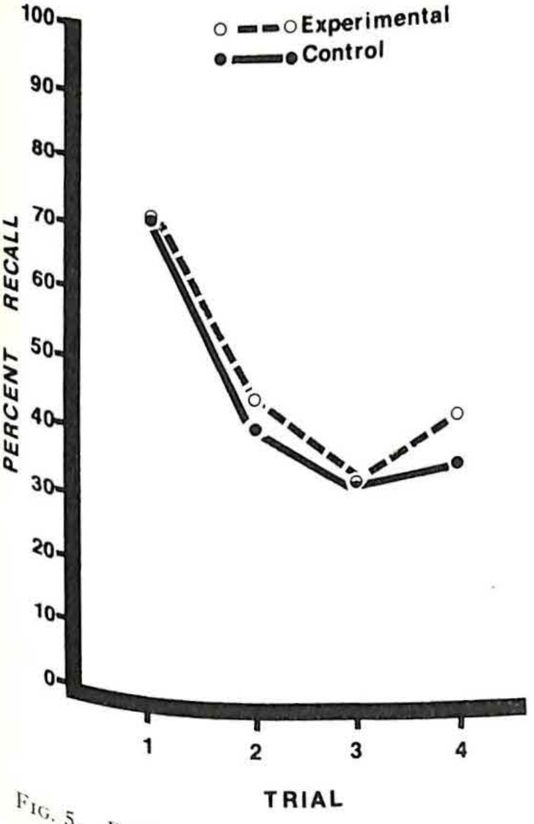


Fig. 5. Performance on the heavy–light shift.

A statement made earlier referred to the encoding in this situation as being automatic and spontaneous, and some justification for this position is in order. In several of the above experiments wherein significant release from PI was obtained, the experimental *S*s were queried after the experiment as to whether or not they noted a difference among the triads. With considerable unanimity the responses were negative. The *S*s just did not seem to be aware of the changing class of material, thereby implying an automatic and spontaneous character to their encoding. This is one reason why an interference type of interpretation is preferred for this phenomenon rather than a perceptual one.[3]

---

[3] This statement is not meant to imply anything about interference in short-term store, since the nature of the present procedure probably removes the data from the same realm as that which produces the recency effect in free recall.

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 1969, **80** (3, Pt. 2).

Bousfield, W. A., & Puff, C. R. Clustering as a function of response dominance. *Journal of Experimental Psychology*, 1964, **67**, 76–79.

BOWER, G. A multicomponent theory of memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* New York: Academic Press, 1967.

GOGGIN, J., & WICKENS, D. D. Proactive interference and language change in STM. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 453–458.

HUDSON, R. L. Category clustering as a function of level of information and number of stimulus presentations. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 1106–1108.

LOESS, H. Short-term memory, word class, and sequence of items. *Journal of Experimental Psychology*, 1967, **74**, 556–561.

MORTON, J. A. A functional model of memory. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970.

NORMAN, D., & RUMELHART, D. E. A system for perception and memory. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

RICHARDSON, J. The learning of concept names mediated by concept examples. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 281–288.

UNDERWOOD, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–573.

UNDERWOOD, B. J., & RICHARDSON, J. Some verbal materials for the study of concept formation. *Psychological Bulletin*, 1956, **53**, 84–95. (a)

UNDERWOOD, B. J., & RICHARDSON, J. Verbal concept learning as a function of instructions and dominance level. *Journal of Experimental Psychology*, 1956, **51**, 229–238. (b)

WICKENS, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

(Received March 2, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 307–314

INFORMATION-PROCESSING ANALYSIS OF VERBAL LEARNING[1]

WILLIAM A. JOHNSTON,[2] ROLLIE R. WAGSTAFF, AND DOUGLAS GRIFFITH

University of Utah

The amount of *S*'s processing capacity expended on a verbal learning task was inferred from his performance of a subsidiary reaction time task: the shorter *S*'s reaction time, the less his expended processing capacity (EPC). In Exp. I, both list organization and free-recall learning reduced EPC, but the learning effect was restricted to EPC during list study. In Exp. II, special procedures ensured that *S* had to process the list on every study trial, but had to recall only one word on every recall trial. Learning emerged in terms of EPC during both list study and word recall. It was concluded that learning and list organization facilitate both encoding and retrieval.

Viewed as an enhancement in information flow, learning should show up on at least three dimensions of information processing: speed, accuracy, and ease. As learning progresses, the relevant information should be processed more quickly, accurately, and easily. Processing ease can be conceptualized in terms of system capacity consumed: the easier a task, the less the processing capacity required to perform it. While learning frequently has been demonstrated in terms of speed and accuracy, the effect of practice on processing ease has not thoroughly been tested. The present research sought not only to afford this test, but also to elucidate the processing loci of learning. Specifically, is learning a phenomenon of input processing, output processing, or both?

The measure of processing ease employed in the present work will henceforth be referred to as *expended processing capacity* (EPC). The measurement of EPC requires the use of a divided attention methodology in which *S* performs a primary task and subsidiary task concurrently. Assuming that both tasks compete for the limited capacity of the same central processor, performance of the subsidiary task should indicate how much of this capacity is consumed by the primary task. The more capacity expended on the primary task, the less residual capacity for the subsidiary task, and the poorer the performance of the subsidiary task should be. Several studies attest to the feasibility and utility of the EPC measure (Johnston, Greenberg, Fisher, & Martin, 1970; Johnston, Griffith, & Wagstaff, 1972; Posner & Boies, 1971; Trumbo & Milone, 1971).

In the present experiments, a verbal learning task served the primary role, and a simple reaction time (RT) task served the subsidiary role. As *S* performed the verbal learning task, he also monitored a light and pushed a button whenever he saw the light brighten. Hence, EPC for the verbal learning task was measured as RT to light signals. The fundamental hypothesis tested was that EPC should be an inverse function of learning trials; as the processing capacity consumed by the verbal learning task decreases as a result of learning, more capacity will be available for the RT task and RT should decrease. By programming light signals during both the presentation and the recall phases of the learning trials, it was possible to assess the processing loci of learning. If learning is a phenomenon of input processing, then EPC during list study should decline across trials; if learning is an output or retrieval phenomenon, then the decline in EPC should show up during recall.

EXPERIMENT I

The first study employed a free-recall paradigm: Nine-word lists were learned for 12 trials each, where a trial consisted of a list presentation phase followed by a free-recall phase. Additional variables included list organization and signal intensity. List organization was manipulated both to assess the effect of that variable on EPC and to establish the generality of any learning effects on EPC. Signal intensity was varied for methodological reasons primarily—to determine if one level of signal intensity is more conducive than the other to the measurement of EPC. Finally, control conditions were implemented in which the primary and subsidiary tasks were performed separately.

---

[1] This research was supported by Grant OEG-8-70-0200 (508) from the United States Office of Education.

[2] Requests for reprints should be sent to William A. Johnston, Department of Psychology, University of Utah, Salt Lake City, Utah 84112.

Method

Subjects and design.—The *S*s were 12 male undergraduates recruited from upper-division psychology courses. The *S*s served for six 45-min. sessions at the rate of one session a day. Each *S* received $1.50 per session plus incentive pay; the latter was determined by a system described below.

The design entailed eight basic treatments: two RT-only conditions (low vs. high signal intensity), two verbal-only conditions (categorized vs. noncategorized lists), and four divided-attention or dual-task conditions (the two levels of signal intensity factorially combined with the two levels of list organization). Phase (presentation vs. recall) and trials were manipulated in all verbal-only and dual-task conditions, and nominal analogs of these variables were represented in the RT-only conditions.

Tasks and apparatus.—The verbal-only task involved an aural presentation of the same nine words, in different orders, for 12 trials. The words were spoken in a male voice at the rate of one word every second, and each presentation of the list was followed by a 9-sec. free-recall period. The prerecorded word lists were played to *S* via a stereo tape recorder and headsets, and *S*'s spoken free recall was tape recorded. The presentation and recall phases of a trial were introduced by "ready start" and "recall" commands, respectively. The entire sequence of 12 trials was terminated by a "stop" command, and was about 252 sec. in duration. The RT-only task was temporally equivalent to the verbal-only task, but "phases" and "trials" were not demarcated for *S*. That is, upon a "ready start" command, *S* merely monitored the stimulus light and detected signals until the "stop" command was issued some 252 sec. later. Under dual-task conditions, *S* simply performed the verbal and RT tasks concurrently for 252 sec. Each 252-sec. segment of a task will henceforth be identified as a *trial block*.

The full details of the apparatus are provided in another report (Johnston et al., 1972). With his preferred finger resting on the detection button, *S* faced a .25-in. light source mounted at approximately eye level. The standard brightness of the light was 8 ftl., and the signal brightness was 32 and 64 ftl. for low- and high-intensity signals, respectively. The 100-msec. increase in brightness appeared subjectively as a barely detectable flicker for low-intensity signals but as a noticeable flash for high-intensity signals. The light signals were preprogrammed to occur at the desired temporal points. An electronic timer started in coincidence with a signal and was stopped by *S*'s button push. If *S* failed to respond within 2 sec., *E* stopped the timer himself and recorded a *miss* for *S*.

Procedure.—The first two sessions were devoted to practice, and the last four to the experiment proper. General instructions and four trial blocks each for RT only and verbal only were provided on the first session. The second session consisted of six dual-task trial blocks bounded at the beginning and end by an RT-only trial block. Only low signal intensity and noncategorized lists were used during practice.

Each experimental session was comprised of eight trial blocks: four dual task, two RT only, and two verbal only. The eight trial blocks were administered to each *S* in a unique random order. A 30-sec. break was provided between successive trial blocks, and each block was preceded by the identification "dual task," "reaction time only," or "verbal only," as appropriate. List organization and signal intensity were held constant for a given *S* on a given experimental session. However, each combination of these variables was represented once for each *S* across sessions, and three times for each session across *S*s. Overall, each *S* performed four trial blocks under each of the eight basic treatments.

Pursuant to the rationale behind the measurement of EPC, the instructions encouraged *S* to devote as much of his attention as necessary to the verbal component of dual-task trials, and only his "left-over" attention to the RT component. Starting with the second practice session, these instructions were supplemented by a monetary incentive system. Specifically, each *S* could earn a maximum of $.25 incentive pay on each trial block. On dual-task trial blocks, 80% of *S*'s maximum incentive pay was based on his verbal (recall) performance, and only 20% was based on his RT performance. Of course, 100% of *S*'s maximum incentive pay was based on verbal performance and RT performance on verbal-only and RT-only trial blocks, respectively. Altogether, *S*s earned 85% ($8.50) of their maximum possible incentive pay, and 95% of their earnings from dual-task trial blocks was determined by their verbal performance.

List construction.—The list items were drawn from 36 of the categories represented in the Battig and Montague (1969) norms: The categories were chosen randomly with the restriction that the seventh item of each category have a frequency of at least 100. The 90 words required for the 10 noncategorized practice lists were drawn from the top 3 words of each category. The words were assigned at random to the practice lists with the restriction that the 9 words in a list be from different categories.

The 108 words required for the experimental lists were the fifth, sixth, and seventh words of each category. Each word was used twice: once in a categorized list and once in a noncategorized list. Each of the 12 categorized lists was comprised of the 3 words from each of three categories. The assignment of categories to lists was random without replacement. Each of the 12 noncategorized lists was comprised of 1 word from each of nine categories. Within these restrictions, the words were assigned randomly without replacement to the noncategorized lists. Twelve different word-presentation orders were generated randomly for each list.

Signal programming.—Thirty-six signals occurred in each trial block at the rate of one or two signals in each phase (presentation and recall) of every trial. To allow *E* time to record RT and to reset the equipment, the minimum interval between successive signals was 4 sec. Over the entire block of 12 trials, 18 signals occurred in each phase, and 2 signals occurred at each of 9 sec. comprising each phase. During list presentation, signals were programmed to occur directly after words. This was done with the intention of measuring EPC while a word was undergoing advanced processing. Over the four occasions that *S* experienced a given treatment, all signal loci were equally represented in every stage of practice (set of three consecutive trials).

In the event that two signals occurred in a given phase, *S*'s RT was scored for only one signal. One-third of all signals were dummy signals for which RTs were not scored; they were included in the interest of lowering signal predictability. The signals for which RTs were scored were represented as equally as possible at all combinations of phase, signal locus, and practice. The temporal arrangement of the signals in a trial block for one treatment was matched in a trial block for every other treatment, including RT-only treatments. However, this temporal consistency in RT signals had no apparent impact on performance: Overall RT performance was virtually constant across the four experimental sessions.

Results

Recall accuracy.—The mean number of words correctly recalled was examined as a function of three variables: attention condition (verbal only, dual-task low signal intensity, and dual-task high signal intensity), list organization, and trials. An analysis of variance of these data disclosed list organization, F (1, 11) = 48.15, $p < .001$; trials, F (11, 121) = 107.08, $p < .001$; and List Organization × Trials, F (11, 121) = 4.41, $p < .001$, as the only significant effects. Recall was a direct function of both list organization and trials, and the effect of list organization declined across trials (as near perfect recall was approached for both levels of organization). It is especially important to note that recall accuracy was not affected by attention condition: Recall was just as high under the dual-task conditions as under verbal only. The fact that the RT component of the dual-task conditions did not impair recall performance is a strong indication that the measures taken to induce *S* to emphasize the verbal component were successful.

EPC.—False alarms were virtually nonexistent, and the probability of a missed signal varied from .01 to .07 across the six basic treatments involving the RT task. The basic EPC measure computed for *S* was the median of the four RTs representing a particular intersection of conditions. On the assumption that missed signals arise from extreme inattention to the RT task, they were counted as high RTs in the computation of medians.

The primary analysis of EPC involved four variables: attention condition (RT only, dual-task categorized lists, and dual-task noncategorized lists), signal intensity, phase, and trials. All four main effects attained significance in an analysis of variance of these data: attention condition, F (2, 22) = 57.61, $p < .001$; signal intensity, F (1, 11) = 61.65, $p < .001$; phase, F (1, 11) = 18.90, $p < .001$; and trials F (11, 121) = 1.94, $p < .05$. As anticipated, RT was an inverse function of both signal intensity and trials, greater under dual task than RT only, and greater during free recall than list presentation.

While a Newman-Keuls test applied to the main effect of attention condition revealed that mean RT was less in RT only than in either dual-task condition ($p < .05$), the two dual-task conditions were not significantly different. However, significant effects of list organization were uncovered in Newman-Keuls analyses of two interactions: the significant Attention Condition × Phase interaction, F (2, 22) = 9.48, $p < .001$, and the nearly significant Attention Condition × Signal Intensity × Phase interaction, F (2, 22) = 3.11, $p < .10$. The former analysis revealed that categorized lists required less EPC than noncategorized

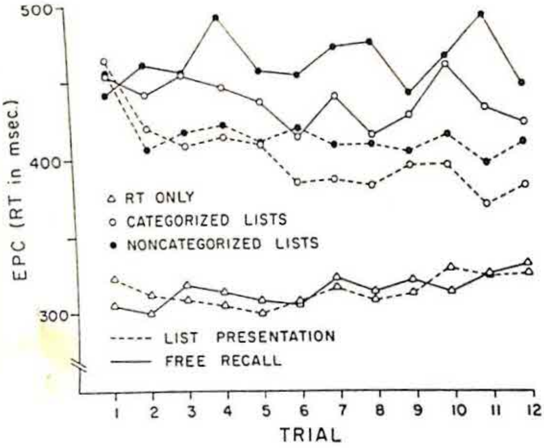


FIG. 1. Expended processing capacity (EPC) learning functions in Exp. I. (The upper four curves represent dual-task conditions, and the lower two curves represent RT-only base lines.)

lists during free recall, and the latter revealed an advantage of categorized lists even during list presentation when signal intensity was low. Hence, list organization enhanced information processing during both phases of a trial, but more so during recall than list presentation.

The other interactions attaining significance were Signal Intensity × Phase, $F\ (1, 11) = 7.84, p < .05$; Attention Condition × Trials, $F\ (22, 242) = 1.84, p < .05$; and Phase × Trials, $F\ (11, 121) = 2.76, p < .005$. The first of these interactions was based on a greater effect of phase under low signal intensity than high, and the bases of the other two can be seen in Fig. 1. Specifically, the effect of trials was confined to the presentation phases of the dual-task conditions.

By collapsing across trials, it was also possible to examine EPC as a function of signal locus, each EPC score being based on five or six RTs. An analysis of variance of these data disclosed the following significant effects involving signal locus: signal locus, $F\ (8, 88) = 3.05, p < .005$; Phase × Signal Locus, $F\ (8, 88) = 2.31, p < .05$; and Attention Condition × Phase × Signal Locus, $F\ (16, 176) = 1.71, p < .05$. All other significant sources of variance were consistent with those obtained in the analysis of trials. The bases of the signal locus effects are depicted in Fig. 2. Specifically, RT was not affected by signal locus in the case of RT only, declined across loci during free recall, and was a **U**-shaped function of loci during list presentation. Whereas only the linear component of the recall function was significant, $F\ (1, 118) = 9.28, p < .01$, the quadratic component attained significance for the list presentation function, $F\ (1, 118) = 15.10, p < .001$.

Discussion

The decrease in EPC across signal loci during free recall undoubtedly was based on a corresponding increase in the likelihood that *S* would terminate his recall attempt and be able to devote his full attention to the RT task. The bowed function obtained for list presentation indicates that *S*s tended to devote more processing capacity to the ends of a list than to the middle. This serial position effect on EPC has been obtained before (Johnston et al., 1972) and suggests the interesting possibility that the classic bowed serial position function in free recall may be due in part to the amount of attention paid to the words during list presentation (Feigenbaum & Simon, 1962).

With respect to list organization, the results indicate that organization influences both input processing and output processing. However, the effect on input processing was less robust, obtaining only under low signal intensity. The tendency for list organization to affect EPC more during free recall than during list pre-

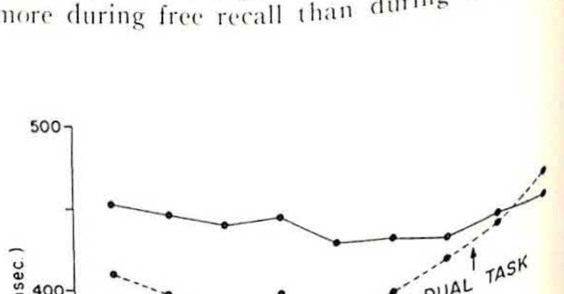


FIG. 2. Expended processing capacity (EPC) as a function of signal locus and phase for dual-task and RT-only conditions in Exp. I.

sentation corroborates prior results (Martin, 1970). Apparently, the well-established effect of list organization of free-recall accuracy is attributable primarily to the processing incurred during retrieval.

The principal objective of Exp. I was to illuminate the information-processing bases of verbal learning. The results are straightforward: EPC did decline over trials, but only during list presentation. This basic finding appears to have some generality as it held for both levels of list organization and for both levels of signal intensity. However, two possible artifacts may render the finding misleading. First, the failure for learning to show up in terms of EPC during recall may have resulted from offsetting effects: a decline in the processing required to retrieve each word counteracted by an increase in the number of words retrieved. Second, the learning that did show up in terms of EPC during list presentation may have resulted from a tendency for *S* to simply ignore a list once it was learned. Hence, learning may not so much enhance input processing as to eliminate the need for it. Experiment II sought to examine the effect of practice on EPC under conditions designed to remove the possible artifacts in Exp. I.

EXPERIMENT II

A probed-recall learning task was used in which *S* had to recall only one of six list words on each trial. The target word (e.g., GOLF) on a given trial was probed for by the category name (e.g., SPORT). The fact that *S* had to recall only one word per trial should have eliminated the first possible artifact referred to above with respect to Exp. I. Any decrease in EPC during word retrieval could not have been masked by an increase in the number of words retrieved. In order to eliminate the second possible artifact in Exp. I, an intrusion-monitoring task was added to the study phases of dual-task trials. Specifically, in addition to studying a list for verbal learning purposes, *S* had to monitor the list in order to detect intrusions. An intrusion was the replacement of a list word by another word from the same category (e.g., GOLF replaced by BASKETBALL). The intrusion-monitoring task prohibited *S* from ignoring a list once it was learned.

Method

Subjects and design.—The *S*s were 24 male undergraduates recruited from introductory psychology courses for extra credit and incentive pay. The basic design was implemented under dual-task conditions and was a Phase (presentation and recall) × Trials (1–18) × Serial Position (1–6) factorial with repeated measures on all factors. In addition, RT-only trials were included to establish a base line for assessing EPC on dual-task trials.

Tasks and apparatus.—The verbal component of dual-task trials involved the aural presentation of six words for 18 trials; on six of the trials, one of the words was replaced by an intrusion. The words were presented at the rate of one word every 2 sec., and 3 sec. were allowed for probed recall. The presentation phase of a trial was introduced by "ready start," and the recall phase was demarcated by a brief click followed by the category probe. A category probe contained no more than three syllables and denoted the category to which the target word belonged. The entire sequence of 18 trials was about 4 min. in duration. The RT component of dual-task trials consisted of low-intensity light signals calling for detection responses as in Exp. I.

The apparatus was the same as for Exp. I except for the modifications required by the intrusion-detection task. An electronic timer started in coincidence with an intrusion and was stopped by *S*'s detection response. An *S* detected intrusions by pushing a button with his left index finger and detected light signals by pushing a separate button with his right index finger.

Procedure.—The *S*s served for four 40-min. sessions at the rate of one session a day. The first session was devoted to practice on the RT task by itself, the verbal learning task (including intrusion monitoring) by itself, and then the two tasks concurrently. In addition, *S*s were familiarized with the 36 category probes used in the experiment. The experiment proper was conducted on the last three sessions.

Each experimental session consisted of a succession of four dual-task trial blocks bounded at the beginning and end by a 2-min. segment of RT only. A different six-word list was learned on each trial block, yielding a total of 12 lists over the three experimental sessions. The 12 lists were administered in four different orders, six *S*s to an order. A 2-min. break was provided after the second list in each session, and 30-sec. breaks were provided after each of the other three lists as well as after the initial segment of RT only.

An incentive system was again used to encourage *S* to emphasize the verbal component of dual-task trials. Errors in intrusion detections and recall contributed equally to the incentive pay. Moreover, in order to ensure a conservative base line for inferring EPC, no incentive pay was provided for RT-only performance.

List construction and signal programming.—The 72 words required for the experimental lists were the

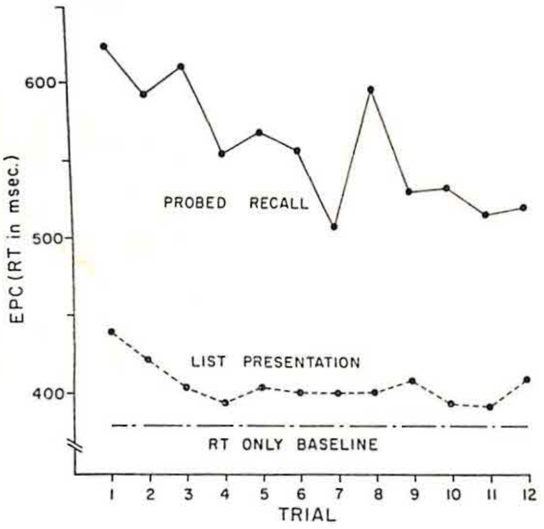


FIG. 3. Expended processing capacity (EPC) learning functions for list presentation and probed recall in Exp. II.

sixth and seventh words from the 36 categories used in Exp. I. The words were assigned at random to the 12 lists with the restriction that the 2 words from the same category be assigned to different lists. The 6 words of each list appeared in different orders across the 18 trials. Each word was replaced by an intrusion once, and the 72 intrusion words required were the second and third words of each category. So that *S* would be sufficiently familiar with a list to detect intrusions, intrusions never occurred on the first 4 trials (and *S* was so informed). The 6 trials containing the intrusions were assigned to Trials 5–18 at random with the restrictions that no more than 2 intrusion trials occur in a row and intrusion trials be assigned as evenly as possible to the 14 permissible trial slots across the 12 lists.

The RT signals occurred at varying intersignal intervals which averaged approximately 6 sec. From one to three signals occurred during each presentation trial, and one occurred on two-thirds of the probed-recall intervals in a trial block. Half of the signals were dummy signals serving signal expectancy purposes, and half were critical signals for which RTs were measured. Critical signals occurred directly after list words or category probes, while dummy signals occurred at later positions relative to words and probes.

The precise composition of each dual-task trial was constrained by counterbalancing procedures too numerous to specify. Suffice it to say that every effort was made to control all relevant factors, such as the serial position containing the target word on a given trial, the serial positions accompanied by critical and dummy RT signals, the specific list words serving as target words, etc. The net result was that each combination of serial position and trial (discounting intrusion trials) was represented by two critical RT signals in the case of list presentation, and one critical signal in the case of probed recall. On intrusion trials there was no overlap in the words probed for, accompanied by RT signals, and replaced by intrusions. The main analyses were conducted on the data for the 12 nonintrusion trials. These 12 trials are distinguished on an ordinal basis only; across the 12 lists, nonintrusion trials were represented in all 18 trial positions.

Results

Verbal performance.—Recall accuracy was high from the start: 80% correct on Trial 1 and 96% by Trial 4. Mean recall latency was determined for eight of the *S*s, averaging 1,180 msec. overall, and declining from 1,406 to 1,103 msec. across the 12 nonintrusion trials, $F\ (11,\ 77) = 7.34$, $p < .001$. The latency function achieved asymptote by about the fifth trial. Finally, intrusion-detection accuracy was 97%, and the latency of these detections averaged 814 msec. Hence, the intrusion-monitoring requirement apparently did force *S* to pay attention to list presentation throughout learning.

EPC.—False alarms were again rare and only around 6% of the RT signals were missed. The analyses of EPC were confined to critical RTs from the 12 nonintrusion trials A missed signal was assigned the value of the longest RT yielded by any of the 24 *S*s for that particular intersection of phase, trials, and serial position. In order to exclude extremely long RTs, all RTs were restricted to a maximum value of 999 msec. Mean RTs were 380, 406, and 558 msec. for RT only, list presentation, and probed recall, respectively. Thus, both list study and probed recall required processing capacity, and EPC was greater for recall than for list study.

Each *S* yielded two RTs for each combination of nonintrusion trial and serial position during list presentation, and the mean of these two RTs was the basic unit of analysis. All three experimental sources of variance attained significance in an analysis of variance: trials, $F\ (11,\ 253) = 4.37$, $p < .001$; serial position, $F\ (5, 115) = 3.78$, $p < .005$; and Trials × Serial Position, $F\ (55, 1265) = 1.58$, $p < .005$. The effect of trials is shown in Fig. 3: EPC during list presentation declined across trials, attain ing asymptote at about Trial 4. The serial position effect was much like that found in Exp. I, only more attenuated. The interaction was based on an erratic tendency for trials effect to hold more for some serial positions than others. For some inexplicable reason, the effect of trials was absent at Serial Position 4, but yet was evident at all other positions.

Each *S* yielded only one RT for each combination of trials and serial (target) position at probed recall. Only trials, F (11, 253) = 8.04, $p < .001$, and Trials × Serial Position, F (55, 1,265) = 2.65, $p < .001$, attained significance in an analysis of variance. The effect of trials is evident in Fig. 3: EPC during probed recall declined across nonintrusion trials. Except for the aberrant jump in EPC on Trial 8, this decline was quite systematic. The decline was robust as well, holding up across all serial positions. The only discernible basis of the interaction was that EPC tended to be an inverted **U**-shaped function of serial position until around Trial 5, at which point the function flattened out.

Discussion

The results indicate that the absence of a learning effect on EPC during recall in Exp. I was due to opposing factors: a decrease in EPC per word recalled, masked by an increase in the number of words recalled. The latter factor was controlled in Exp. II, and a learning effect did emerge in terms of EPC at recall. Apparently, information becomes relatively more accessible to recall as learning progresses. On the other hand, the results suggest that the presence of a learning effect on EPC during list presentation in Exp. I was genuine. Despite the intrusion-monitoring requirement, EPC during list study remained uniformly low from the fifth nonintrusion trial on in Exp. II. If the decrease in EPC across the first four trials were simply a result of inattention to list input, then EPC should have returned to Trial 1 level when attention to input was required by the intrusion-monitoring task on Trial 5. Hence, verbal learning appears to be a phenomenon of both input processing and output processing.

GENERAL DISCUSSION

The principal objective of the present research was to explore verbal learning in terms of ease of information processing. Viewed collectively, the data indicate that practice does reduce EPC, both during list study and word recall. List study may be conceptualized as comprising the retrieval (for perceptual purposes) and manipulation of the memory representations of list words. The retrieval aspect of list study is probably trivial owing to the powerful retrieval cues afforded by the words themselves. Hence, the effect of practice on input processing likely was realized in the manipulation of perceptual information. Learning apparently yields a reduction in the processing devoted to rehearsing, elaborating, organizing, and otherwise systematically encoding list information.

Recall may be conceptualized as comprising the retrieval of the stored representations of list words and the transformation of these representations into overt recall. The retrieval aspect of recall is probably nontrivial owing to relatively ineffective retrieval cues and is likely the main locus of the practice effect. Practice may reduce EPC at word recall by enhancing the accessibility of list information in storage. Just how practice enhances information accessibility is open for speculation. It is tempting to attribute the effects of practice on both list study and word recall to list organization processes. However, the tendency for practice and list organization to have independent effects on EPC in Exp. I mitigates this interpretation. While learning and organizational processes may be intertwined in some cases (Tulving, 1968), they apparently were independent in the present work.

In line with numerous other experiments, EPC was consistently higher during recall than during list presentation (Johnston et al., 1970; Johnston et al., 1972; Martin, 1970; Trumbo & Milone, 1971). Apparently, the processing incurred during recall in standard verbal learning and memory tasks is more demanding than the processing incurred during list study. While this relationship may hold for standard tasks, it should be potentially reversible by using tasks which place relatively heavier demands on input processing. Finally, relative to the RT-only base lines, EPC at recall was substantially higher in Exp. II than in Exp. I. This difference undoubtedly reflects the change from a free-recall procedure in Exp. I to a probed-recall procedure in Exp. II. The notion that EPC is a direct function of the degree of restriction placed on recall (e.g., free vs. probed) has both intuitive appeal and empirical support (Johnston et al., 1970, Exp. III and IV).

REFERENCES

BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 1969, **80** (3, Pt. 2).

FEIGENBAUM, E. A., & SIMON, H. A theory of the serial position effect. *British Journal of Psychology*, 1962, **53**, 307–320.

JOHNSTON, W. A., GREENBERG, S. N., FISHER, R. P., & MARTIN, D. W. Divided attention: A vehicle for monitoring memory processes. *Journal of Experimental Psychology*, 1970, **83**, 164–171.

JOHNSTON, W. A., GRIFFITH, D., & WAGSTAFF, R. R. Speed, accuracy, and ease of recall. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 512–520.

MARTIN, D. W. Residual processing capacity during verbal organization in memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 391–397.

POSNER, M. I., & BOIES, S. J. Components of attention. *Psychological Review*, 1971, **78**, 391–408.

TRUMBO, D., & MILONE, F. Primary task performance as a function of encoding, retention, and recall in a secondary task. *Journal of Experimental Psychology*, 1971, **91**, 273–279.

TULVING, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N.J.: Prentice-Hall, 1968.

(Received March 4, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 315-320

MOVEMENT VELOCITY AND MOVEMENT TIME AS DETERMINERS OF DEGREE OF PREPROGRAMMING IN SIMPLE MOVEMENTS

RICHARD A. SCHMIDT[1] AND DAVID G. RUSSELL

Department of Physical Education, University of Michigan

Twelve *S*s performed a timing task involving a striking movement in order to estimate the degree of feedback control in discrete movements. The *S*s were pretrained to make 22.8- or 49.5-cm. movements in exactly 150 or 750 msec., and *S*s used this movement later in the timing task. A 2 × 2 repeated-measures design (Movement Time [MT] × Movement Distance) was used, and the index of preprogramming (IP) was used to assess the degree of feedback involvement. Reducing the MT from 750 to 150 msec. nearly doubled the IP implying greatly reduced feedback control in the 150-msec. movement, but doubling the movement velocity by lengthening the movement distance (MT constant) had no effect on the IP. This supported the notion that lack of feedback involvement in movement of short duration (e.g., 150 msec.) is due to a temporal limitation in feedback processing, and that even very slow responses with 150-msec. MTs may be highly preprogrammed.

Recently there has been considerable interest (e.g., Keele, 1968; Schmidt, 1969, 1972) in understanding the development and operation of the "motor program" defined by Keele as a sequence of stored commands that is "structured before the movement begins and allows the entire sequence to be carried out uninfluenced by peripheral feedback [p. 387]." Support for the notion of the motor program is based upon the inability of feedback control mechanisms to account for the execution of controlled movement in a wide variety of situations. For example, Lashley (1917) using a deafferented human and Taub and Berman (1968) and Wilson (1961) using deafferented monkeys and locusts, respectively, found that patterned movement could be performed in the absence of peripheral feedback control. In addition, since the feedback processing time is on the order of 120 msec. for kinesthetic (Chernikoff & Taylor, 1952) and 200 msec. for visual (Keele & Posner, 1968) information, the fact that *S*s can easily initiate, carry out, and terminate a rapid manual response within 100 msec. seems to imply that the instructions to stop the movement must be structured prior to the beginning of the movement.

The recent concern for the motor program has naturally led to a search for the laws which determine how it is acquired and which govern its operation. One consistent conclusion depends upon the long delays in processing feedback and states that the length of the movement time (MT) for the response would be a major determiner of the degree of programming. Evidence for this view has come from findings that movement cannot be interrupted until it had run its course for approximately 200 msec. (Henry, 1960; Henry & Harrison, 1961), and that longer movements enable feedback to provide a basis for movement control (e.g., Ellis, Schmidt, & Wade, 1968). Therefore, it seems fairly certain that MT is critical in determining the degree to which movement is preprogrammed.

Schmidt (1969, 1972) has studied other variables which determine the extent of movement programming using the index of preprogramming (IP). Faster movement speed (distance held constant), shorter movement distance, and higher inertial loading all increased the IP, implying decreased reliance upon feedback and increased dependence upon the motor program control. However, the above manipulation of speed, distance, and inertial load are all confounded with MT. For example, increasing the movement speed (movement distance held constant) decreased the MT, and one cannot be sure that the speed effects on the IP were due to movement speed per se or to the secondary effects on MT; the same can be said of movement distance and load. While Henry (1960; Henry & Harrison, 1961) has suggested that a movement must be made at maximal speed in order to be programmed and that decreased movement velocity would result in increased utilization of feedback, there is no theoretical reason why this must be so. In fact, it is possible that the only variable which determines programming is MT, and that a very slow 150-msec. response is as free of feedback control as a fast 150-msec. response. The present investigation tested this hypothesis by manipulating movement distance (MT held constant) to vary movement velocity, with the degree of preprogramming in the movements measured by the IP as described by Schmidt (1972).

---

[1] Requests for reprints should be sent to Richard A. Schmidt, Department of Physical Education, Barbour Gymnasium, University of Michigan, Ann Arbor, Michigan 48104.

METHOD

Subjects.—The *S*s were 12 (8 male and 4 female) right-handed undergraduate and graduate student volunteers from University of Michigan. They were not paid for their services, and all were naive to the purposes of the study.

Apparatus.—A microswitch (*S* switch) and a hinged Masonite target (10 cm. square) were mounted on the top of a large wooden table. The *S* switch and target were mounted along a line parallel to and 9 cm. from the edge of the table, perpendicular to the direction the seated *S* was facing, the *S* switch on the right and target on *S*'s left. The location of the *S* switch was fixed, but the target could be moved to vary the *S* switch–target distance. With the target in the vertical position, a set of contact points was "closed," and the operation of a toggle switch (*E* switch) by *E* started two .01-sec. timers (Standard Type S-1, 6-v. dc clutch, 1-sec. sweep). One of the timers (*S* clock) was placed upon the table in *S*'s full view, the other (*E* clock) being visible only to *E*. Knocking over the target to *S*'s left "opened" the contacts and both clocks stopped. In addition, with the target vertical and *S* switch held down ("open"), releasing the *S* switch started a third clock (MT clock) which could be stopped by knocking over the target. The MT clock could be seen only by *E*.

Task.—With *S*'s right hand depressing the *S* switch, *E* operated the *E* switch to start the *E* and *S* clocks. The *S*'s task was to move from the *S* switch to knock over the target at the moment that the *S* (and *E*) clock reached exactly 2.00 sec. (vertical), and *S* was to do this with an MT of either 150 or 750 msec., depending upon the conditions under which he was practicing. Knocking over the target stopped both the *E* and *S* clocks and enabled *S* to see the error in his response. The major task was to move in such a way that the *S* clock hand stopped exactly at the 2.00-sec. mark.

Design.—The design was a 2×2 arrangement, with the variables being movement distance (either 22.8 or 49.5 cm.) and MT (either 150 or 750 msec.). Varying the movement distance for a given level of MT manipulated the movement velocity, and permitted an evaluation of the effects of velocity on the IP uncontaminated by covariation in MT. The 12 *S*s were used in all 4 treatment combinations, with the order of treatments determined by a 4×4 Latin square. The *S*s were assigned at random to a row of the square with the restriction that each row have 3 *S*s.

Procedure.—Holding the MT constant at 150 or 750 msec. required that pretraining be provided in order that *S*s learn the appropriate movement for each of the four MT–distance pairs. In the first part of each testing session, *S* practiced 50 trials in which he was to learn to move the specified distance in the MT assigned for that day. Knowledge of results about MT (in milliseconds) was provided after each trial. The *S* clock did not operate on pretraining trials.

Following successful completion of these trials, instructions were given concerning the main task which required *S* to knock over the barrier with the just-learned MT so that the *S* clock hand stopped at 2.00 sec. Accuracy in timing the striking of the barrier was clearly emphasized as being more important than moving with a precisely correct MT, but the latter was described as being important as well. The *S*s were instructed to try to correct their responses in progress if they could so as to strike the target at the correct instant. After *E* was sure that *S* understood the instructions, *S* practiced 10 additional MT-alone trials, and then the main timing task was begun.

There were 75 trials run on the main task, with a 15-sec. intertrial interval, except that between Trials 25 and 26 and 50 and 51 a 1-min. rest was provided. The *S* received his own error information from viewing the *S* clock, and to be sure that *S* was attending primarily to the timing of the striking response, he called out either "hit," "too early," or "too late" after each trial. The *E* monitored *S*'s MTs on the MT clock and warned him if his MT was outside of limits set a priori (150 ± 15 msec. or 750 ± 100 msec.), and information about MT was provided only in cases where the MT fell outside of these limits. The testing for a particular MT–distance pair was completed on each day, and this was repeated for a total of 4 days, each with a different MT–distance combination.

Measures recorded.—Two scores were recorded for each trial. Algebraic error (AE) was the difference with respect to sign between the target time of 2.00

sec. and *S*'s arrival time taken from the *E* clock. The *S*, of course, had access to the same information via the *S* clock. MT was recorded from the MT clock and was the time from movement initiation until *S* struck the barrier. A third measure termed starting time (ST) was the instant in the 2.00-sec. period at which *S* began to move, and was calculated by the formula ST = MT − AE. The correlation between ST and AE computed within *S*s over the last 50 trials (transformed to Z') was the IP. The rationale for the IP (see Schmidt, 1972) is that if a movement is totally preprogrammed (feedback produced during the movement having no effect on that movement), then the time of finishing the response (the AE) should be highly correlated with the time it started (the ST). A feedback-controlled response, on the other hand, should permit feedback to adjust the response while it is in progress, and the AE–ST correlation should decrease as the feedback involvement increases. In addition to these logical grounds for validity, the IP changes predictably as a function of experimental variables which should influence the degree programming. Increasing the MT should increase the time available for feedback processing, and should lead to decreased IPs. Schmidt (1969, 1972) found that increased MT created by moving slower or by moving a greater distance decreased the IP as predicted. Also, inertial loading on the movement should permit less effective adjustments in response to feedback relative to unloaded movements, and Schmidt (1969, 1972) has found that increased loading increased the IP. This, plus the fact that transformations known to change the skewness of the ST and AE distributions have no effect on the IP, provides evidence that the IP has reasonable validity as a measure of feedback involvement.

Results

Movement times.—The obtained MTs are given in Table 1 as a function of movement distance and instructed MT. The desired MTs of 150 and 750 msec. were nearly maintained in the timing task, with the mean obtained MTs being 146 msec. for the fast response and 695 msec. for the slow response, these differences being highly significant in a repeated-measures MT × Distance ANOVA, $F\ (1,\ 11) = 176.0$, $p < .001$. There was a tendency for the short distances to produce somewhat shorter MTs, but these differences were not significant, $F\ (1,\ 11) = .9$, $p > .05$. The MT × Distance interaction was not significant, $F\ (1,\ 11) = .3$, $p > .05$. Therefore, *S*s apparently were able to move with approximately the desired MT, even though the MTs were slightly shorter than the a priori values set.

TABLE 1

MEAN OBTAINED MOVEMENT TIMES (MSEC.) AND MOVEMENT SPEEDS UNDER THE FOUR TREATMENT COMBINATIONS

Movement distance (cm.)	Instructed MT (msec.)		
	150	750	$\bar{X}$
22.8	143 (159)	690 (33)	416 (96)
49.5	149 (332)	700 (71)	424 (201)
$\bar{X}$	146 (245)	695 (52)	420 (148)

Note.—MT = movement times; movement speeds in cm/sec are given in parentheses.

TABLE 2

MEAN ABSOLUTE AND ALGEBRAIC ERROR UNDER THE FOUR TREATMENT COMBINATIONS (IN MSEC.)

Movement distance (cm.)	150	750	$\bar{X}$
22.8	50.8 (+31.4)	57.3 (+19.4)	54.1 (+25.5)
49.5	48.2 (+27.6)	52.5 (+21.1)	50.4 (+24.4)
$\bar{X}$	49.5 (+29.6)	54.9 (+20.3)	52.2

Note.—Algebraic errors are given in parentheses. Late responses are positive algebraic errors.

Response accuracy.—The accuracy of responding at 2.00 sec. in the task under the various conditions of distance and MT was evaluated by absolute and algebraic errors, and is presented in Table 2. With respect to algebraic errors, responding under the 150-msec. condition produced more late responding than responding under the 750-msec. condition, and this difference was significant, $F\ (1, 11) = 6.3, p < .05$. However, the distance effects were very small and were not significant, $F\ (1,\ 11) = .1$, $p > .05$. The MT × Distance interaction failed significance as well, $F\ (1,\ 11) = .7$, $p > .05$.

With respect to absolute errors, there was a tendency for slower responding to produce greater inaccuracy for both the

TABLE 3

MEAN INDEXES OF PREPROGRAMMING AND Z' TRANSFORMATIONS OF MOVEMENT TIME (MT)–STARTING TIME (ST) CORRELATIONS UNDER THE FOUR TREATMENT COMBINATIONS

Movement distance (cm.)	Instructed MT (msec.)		$\bar{X}$
	150	750	
22.8	1.270 (.696)	.675 (.702)	.973 (.699)
49.5	1.510 (.563)	.646 (.862)	1.078 (.713)
$\bar{X}$	1.389 (.629)	.660 (.782)	

Note.—Z' transformations of MT–ST correlations are given in parentheses.

short and long distances, but the difference was not significant, F (1, 11) = 1.5, $p >$ .05. Increasing the movement distance produced somewhat greater accuracy, but the differences were not significant, F (1, 11) = 1.6, $p > .05$. Not surprisingly, the MT × Distance interaction failed to reach significance, F (1, 11) = .01, $p > .05$.

Index of preprogramming.—The mean IPs are presented in Table 3. Decreasing the MT from instructed values of 750 to 150 msec. more than doubled the IP from Z' = .66 to 1.39, indicating that the short MT prevented the use of feedback and required the response to be programmed to a greater extent. This difference was significant in a repeated-measures MT × Distance ANOVA, F (1, 11) = 13.6, $p < .01$. However, more than doubling the movement velocity (Table 1) by holding MT constant and increasing movement distance had the effect of increasing the IP slightly for 150-msec. movements and decreasing it slightly for the 750-msec. movements. These distance effects were quite small, however, and were not statistically significant, F (1, 11) = 1.2, $p > .05$. The MT × Distance interaction also failed significance, F (1, 11) = 2.9, $p > .05$. These findings supported the hypothesis that MT would be the only important variable in programming, as holding MT constant and varying velocity by lengthening distance failed to produce reliable effects on the IP.

However, the within-S standard deviation of the MTs more than doubled as the instructed MT increased from 150 to 750 msec., and it was possible that the decreased IP with increasing MT could have been caused by a spurious depressing effect of the elevated MT variability. This argument can be countered in two ways. First, increased variability in MT, if due to "noise" factors entering as the movement progressed, should lead to increased variability of AEs; this did not occur in the present data as evidenced by the nearly constant absolute errors over the two movement times (Table 2). Second, if the increase in MT variability were due to random "noise" factors, then the random error component should not be associated with other events in the trial. However, Table 3 shows that the MT is associated with the ST to an increasing extent as the MT increases, which is contrary to the expectation if the increased variability were due to random "noise." A more likely explanation is that the 150-msec. movements consisted of a programmed segment, while the 750-msec. movement consisted of a programmed initial segment plus a feedback-based correction. Since variabilities of these components should be additive, the result would be the increased MT variability shown. Alternatively, the increased MT could be seen as reflecting the greater number of ways S has to produce the response when the MT is long. In any case, the hypothesis that the decreased IP as a function of increasing MT was caused by "noise" which degraded the AE–ST correlation (the IP) seems untenable.

DISCUSSION

In studies of motor programming conducted earlier by Schmidt (1969, 1972), increasing movement speed, decreasing movement distance, and increasing the inertial load on movement tended to shorten the MT and increase the IP. The finding that IPs increased whenever the MT decreased suggested that MT may be a powerful variable in determining the potential for feedback involvement in movement. This interpretation was not strong, however, as the manipulations of speed, distance, and load were confounded with MT, and it may have been possible for any of these factors to be responsible for the changes in the IPs observed. The present study, therefore, sought to unravel the effects of these variables.

The present findings showed that MT emerged as the most important variable in determining the IP, as decreasing the obtained MT values from 695 to 146 msec. more than doubled the IP, implying that the ST strongly determined the AE in the 150-msec. response, and that *S*s were free to adjust the 750-msec. responses somewhat once they were under way. On the other hand, more than doubling the movement velocity by lengthening the movement distance (holding MT constant) produced no reliable effects on the IP, suggesting that the velocity of movement per se was unimportant in determining the degree of feedback control of movement. This was contrary to the view expressed by Henry (1960; Henry & Harrison, 1961) that a response must be made at maximal speed in order that it be programmed. The present findings suggested that for a given MT, a very slow movement and a very fast movement (with, naturally, differing movement distances) should be equally preprogrammed. As human *S*s make a great number of slow responses in numerous tasks such as tracking, striking, etc., the present findings supported the idea that such responses, even if not at maximal speed, are probably performed under programmed control for at least short periods of time (e.g., 150 msec.).

The findings concerning the measures of timing accuracy were considerably different from those found earlier (Schmidt, 1969), where increasing the MT increased absolute timing errors, especially when the movements were unloaded. The interpretation of those data was that a shorter MT required a shorter interval of time to be estimated in planning the movement. Since it has been found (see Schmidt, 1968) that the accuracy of prediction of the end of an interval decreases with increasing interval lengths, *S*s were thought to be more uncertain about the MT and therefore about when to begin the movement (the ST). Because the movements were programmed, uncertainty about the ST created error in the timing response. Although there was a hint of decreased absolute error with decreased MT in the present data, the differences were quite small and were not significant. While it is possible that the lack of significance was due to the rather small sample size in the present experiment, another explanation is suggested. In the earlier article (Schmidt, 1969), the display was a small moving target. Since increasing the MT required *S* to initiate the movement when the target was more distant, the greater distance and the resultant decreased visual angle for a given amount of target movement may have resulted in a poorer basis about when *S* should begin the movement with respect to the target location. However in the present study, an earlier ST was not associated with poorer discrimination of target location, since the "target" was the hand on the clock face. Thus it could be that, in the earlier study (Schmidt, 1968), increasing error associated with longer MTs was confounded with visual discrimination of the target's location, and this effect disappeared in the present study when visual discrimination was held constant across MTs. The present finding of more late AEs when MT was shortened was not found by Schmidt (1969), and the reasons for this unexpected finding are not clear at present.

The present findings suggested that motor programs may be used extensively when MTs are short, and that the inability of *S*s to use feedback to control movement is due chiefly to a temporal limitation in *S*s' ability to process peripheral information. While it may be true that MT is the only variable which determines *S*'s ability to use feedback in movements, it may also be that other variables such as movement complexity and practice are superimposed on MT.

REFERENCES

CHERNIKOFF, R., & TAYLOR, F. V. Reaction time to kinesthetic stimulation resulting from sudden arm displacement. *Journal of Experimental Psychology*, 1952, **43**, 1–8.

ELLIS, M. J., SCHMIDT, R. A., & WADE, M. G. Proprioception variables as determinants of lapsed-time estimation. *Ergonomics*, 1968, **11**, 577–586.

HENRY, F. M. Increased response latency for complicated movements and a "memory-drum" theory of neuromotor reaction. *Research Quarterly*, 1960, **31**, 448–458.

HENRY, F. M., & HARRISON, J. S. Refractoriness of a fast movement. *Perceptual and Motor Skills*, 1961, **13**, 351–354.

KEELE, S. W. Movement control in skilled motor performance. *Psychological Bulletin*, 1968, **70**, 387–403.

KEELE, S. W., & POSNER, M. I. Processing of feedback in rapid movements. *Journal of Experimental Psychology*, 1968, **77**, 353–363.

LASHLEY, K. S. The accuracy of movement in the absence of excitation from the moving organ. *American Journal of Physiology*, 1917, **43**, 169–194.

SCHMIDT, R. A. Anticipation and timing in human motor performance. *Psychological Bulletin*, 1968, **70**, 631–646.

SCHMIDT, R. A. Movement time as a determiner of timing accuracy. *Journal of Experimental Psychology*, 1969, **79**, 43–47.

SCHMIDT, R. A. The index of preprogramming (IP): A statistical method for evaluating the role of feedback in simple movements. *Psychonomic Science*, 1972, **27**, 83–85.

TAUB, E., & BERMAN, A. J. Movement and learning in the absence of sensory feedback. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior*. Homewood, Ill.: Dorsey, 1968.

WILSON, D. M. The central nervous control of flight in a locust. *Journal of Experimental Biology*, 1961, **38**, 471–490.

(Received March 10, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 321–327

EFFECT OF SLEEP ON MEMORY:

III. CONTROLLING FOR TIME-OF-DAY EFFECTS[1]

TERRY R. BARRETT AND BRUCE R. EKSTRAND[2]

University of Colorado

Retention of a paired-associate list was compared in three groups of *S*s who all learned and recalled at the same time of day (the retention interval was always defined from 2:50 a.m. to 6:50 a.m.). One group was awake during the interval, one group experienced sleep characterized by high amounts of Stage IV sleep and low amounts of REM sleep, and the third group experienced sleep characterized by low amounts of Stage IV and high amounts of REM. The results demonstrated that memory over a sleep interval was superior to memory over a waking interval. Of the two sleep intervals, retention over the high REM sleep interval was inferior to retention over the high Stage IV interval. These two effects cannot be explained as artifacts of a circadian rhythm effect on memory. Implications of the results for understanding the effects of sleep on memory were discussed.

One of the cornerstone phenomena of the interference theory of forgetting is the "sleep effect." Subjects who sleep during a retention interval remember more than *S*s who are awake (Ekstrand, 1967; Jenkins & Dallenbach, 1924; vanOrmer, 1932). According to the interference theory, sleep prevents interpolated learning (IL) during the retention interval which results in a decrease in retroactive inhibition (RI); i.e., there is better recall. Awake *S*s are, of course, presumably engaging in significant amounts of IL and thus suffer from significant amounts of RI; i.e., they remember less than the sleeping *S*s.

Recently, this interference interpretation of the sleep effect has been challenged by Yaroush, Sullivan, and Ekstrand (1971). These investigators demonstrated that retention over the first 4 hr. of a night's sleep was superior to retention over the second 4 hr. which was not superior to an awake condition. In terms of the physiologically defined sleep stages (Rechtschaffen & Kales, 1968), the first half of the night contains most of the Stage IV sleep and very little REM sleep whereas the second half contains most of the REM sleep and very little Stage IV sleep. The Yaroush et al. (1971) data thus strongly suggest the possibility that Stage IV sleep facilitates memory or that Stage REM interferes with memory, or both. Both the first-half and second-half *S*s are sleeping during the retention interval so the differential sleep effect is difficult to explain in terms of an interference theory of the sleep effect, unless one wishes to postulate that dreaming (as defined by REM sleep) results in IL.

To our knowledge, all of the sleep studies that have been done, including the Yaroush et al. (1971) experiment, have confounded the sleep-awake variable with time of day. For obvious reasons of convenience, the sleep *S*s have always been run at night, typically learning late at night and recalling first thing in the morning. The awake *S*s have always been run in the daytime, typically learning in the morning and recalling later in the afternoon. In the Yaroush et al. (1971) experiment, the retention interval for the first-half condition was defined from approximately 11:00 p.m. until 3:00 a.m. and from approximately 3:00 a.m. until 7:00 a.m. for the second-half condition. The awake condition was run during regular, daytime hours. One additional reason for running the awake *S*s during the daytime stems from a consideration of the interference theory. Forgetting in these *S*s is presumably due to extraexperimental IL which *S*s accomplish during the retention interval. The *S*s learn the original task (OL) and then are released from the laboratory to carry on their normal activities. If one were to run the awake *S*s at night at the same time as the sleep *S*s, he would not have a "normal" time period for extraexperimental learning. Another tact would be to attempt to run the sleep *S*s during the daytime at the same time as the awake *S*s. This, however, would probably result in an abnormal sleep pattern. Such a procedure would mean that *S*s were taking naps instead of normal sleep, or would require prior sleep deprivation to insure that *S*s could sleep easily during the daytime. Alternatively, one could use night-shift workers who normally sleep during the daytime.

---

[1] This research was supported by Grant MH-15655 from the National Institute of Mental Health to Bruce R. Ekstrand. This report represents Publication No. 22 of the Institute for the Study of Intellectual Behavior which is supported by the Graduate School of the University of Colorado. We wish to thank all the staff of the Research Center for the Study of Sleep and Dreams, a division of the Institute for the Study of Intellectual Behavior. Without their cooperation and dedication this research would not have been possible.

[2] Requests for reprints should be sent to Bruce R. Ekstrand, Department of Psychology, University of Colorado, Boulder, Colorado 80302.

Considering all the difficulties that one would encounter if he attempted to control for time-of-day effects, it is no wonder that no one has done so. Typically, there have been no differences in the speed of OL at the various times when *S*s are run and so the investigators have argued their way around the confounding by saying there are no time-of-day effects. The main criticism of this approach is that a gross measure such as trials to criterion may not be sensitive to time-of-day effects whereas the presumably more sensitive dependent variable of recall may in fact be influenced by time of day.

It is our feeling that in view of the importance of the sleep experiments for developing a general theory of memory, it is necessary to determine whether or not the sleep effects that have been demonstrated might possibly be artifacts of time of day. This is particularly true in view of the great interest presently being shown in circadian rhythms and other biological rhythms (see Luce, 1970). There is little doubt that these rhythms influence a wide variety of behaviors, and as such, there is reason to believe that the sleep effects already mentioned may have been influenced by the effects of these rhythms. The purpose of the present experiment, therefore, was to compare forgetting over intervals of sleep with forgetting over intervals of wakefulness, when the intervals for both conditions occurred at the same time during the day. We also compared the retention of two different sleep conditions, one with high Stage IV sleep and low REM sleep and the other with high REM sleep and low Stage IV sleep, both run at the same time of day.

METHOD

Design and subjects.—There were three conditions in the experiment, first half (Cond. F), second half (Cond. S), and awake (Cond. A). The retention interval in all three conditions was designed to be 4 hr. in length and to occur always between 2:50 a.m. and 6:50 a.m. For *S*s in Cond. F, this interval represented the first half of a night's sleep; whereas, it was the second half of a night's sleep for *S*s in Cond. S. In Cond. A, *S*s were awake during this interval. Each *S* learned a paired-associate list consisting of 15 unrelated pairs of high-frequency words and was tested for recall of the list 4 hr. later. There were 10 *S*s in each of the three conditions. The *Ss* were volunteers from introductory psychology classes at the University of Colorado, serving to fulfill a course requirement. All *S*s volunteered to spend 1 night sleeping in the laboratory, after being told they might lose some sleep. Each *S* was then randomly assigned to serve in one of the three conditions.

Procedure.—The learning and recall procedures were identical in all three conditions. Each *S* learned one PA list by the study-test method to a criterion of 10 out of 15 correct responses on one trial. The presentation rate was 2 sec. during both study and test portions and the intertrial interval was also 2 sec. Learning was carried out on a Stowe memory drum. Six different random orders (three study and three test) of the list were used to prevent serial learning. After completion of OL, each *S* was read a set of antirehearsal instructions. Four hours later *S*s were tested for recall of the list. In order to make sure that all *S*s were fully awake at the time of learning and recall, they were required to perform a mirror star-tracing task just before learning and again just before recall. Each *S* traced one star with his preferred hand followed by one star with his nonpreferred hand. This tracing task was used in all conditions. At the time of recall, following the mirror-tracing task, *S* was given two consecutive paced test trials on the list he had learned. This was followed by an unpaced test

The *S* was given a sheet of paper listing all the stimulus words and was given unlimited time to recall as many of the appropriate responses as possible. This was followed by a matching test. The *S* was given a sheet, randomly listing the stimuli in one column and the responses in another, and was required to match them.

For *Ss* in Cond. F, the retention interval (2:50 a.m. to 6:50 a.m.) was designed to occur during the first half of a night's sleep, containing high amounts of Stage IV sleep and low amounts of Stage REM. These *Ss* arrived at the laboratory about 12:30 a.m. They had been instructed not to sleep at any time during the day prior to the experiment. They stayed awake in the laboratory reading or talking with *Es* until 30 min. before the scheduled time of OL. During this time they also were prepared for physiological monitoring of their sleep. At about 2:00 a.m. these *Ss* went to bed. Thirty minutes later they were "awakened" and completed OL. Following completion of the OL task, these *Ss* returned to bed for the retention interval. The mean time of "lights out" for these *Ss* was exactly 2:50 a.m. The *Ss* then slept for approximately 4 hr. while *E* monitored EEG, EOG, and EMG information in order to assess stages of sleep. They were then awakened for recall which took place at a mean time of 6:52 a.m.

In Cond. S, the retention interval was designed to occur during the second half of a night's sleep, containing high amounts of Stage REM and low amounts of Stage IV. These *Ss* arrived at the laboratory about 10:00 p.m. and were immediately prepared for physiological monitoring. They then went to bed and slept for approximately 4 hr. They were awakened about 2:30 a.m. and completed the OL task, after which they returned to bed (mean lights out was 2:52 a.m.) for the 4-hr. retention interval. Again *E* monitored EEG, EOG, and EMG during the retention interval. This was followed by recall which took place at a mean time of 6:52 a.m. Thus, for Cond. S, the retention interval occurred over the second half of an 8-hr. night of sleep.

In Cond. A, *Ss* were awake during the retention interval. These *Ss* appeared at the laboratory about 10:00 p.m. and were prepared for EEG, EOG, and EMG monitoring. They then retired for a 4-hr. sleep interval, being awakened at about 2:30 a.m. for completion of the OL task. Following the completion of OL, these *Ss* were required to lie down in a room for 15 min. without sleeping. This was used to provide them with a 15-min. period after OL, comparable to the interval the sleep *Ss* have while they are attempting to fall asleep. Thus, the sleep and awake conditions had comparable intervals for rehearsal immediately after OL. After this brief rest interval, the awake *Ss* got up and spent the rest of the retention interval awake in the laboratory. The mean time for the starting of the retention interval was 2:49 a.m. for these awake *Ss*. Approximately 4 hr. later (mean recall time 6:48 a.m.), they were tested for OL recall. During the awake interval, *Ss* were either "entertained" by *Es* in our monitoring room or they read in an adjacent, brightly lit room. We provided magazines and paperbacks and *Ss* were allowed to bring their own reading materials. The typical awake *S* spent some time reading and some time observing the physiological monitoring of other *Ss* sleeping in the laboratory. The *Es* also made a point of engaging *Ss* in conversations about sleep and dream research, attempting to inform them about the modern knowledge of the sleep-dream process. While *Ss* were reading in the adjacent room, *Es* checked on them quite frequently to make sure they were awake.

The *Ss* in the two sleep conditions (F and S) were physiologically monitored during the retention interval. We recorded EEG, EOG, and EMG potentials. The electrode placements, monitoring techniques, and sleep-stage scoring procedures were according to the standard manual in sleep physiology (Rechtschaffen & Kales, 1968). Awake *Ss* were monitored for the 4 hr. prior to the retention interval, while they were sleeping, but not during the retention interval.

RESULTS

Learning.—The means for trials to criterion in the three conditions are shown in Table 1. An analysis of variance revealed no significant effect of condition, $F\ (2,\ 27) < 1$. Table 1 also shows the mean number correct on the criterion trial of OL. Again an analysis of variance revealed no differences among the conditions, $F\ (2,\ 27) < 1$. Based upon these measures it does not appear that the three conditions resulted in different speeds of learning or different final degrees of learning. Thus, the slight sleep deprivation which occurred in Cond. F did not seem to affect learning significantly. Thus differences among conditions in recall cannot be attributed to differences in degree of original learning.

Recall.—Table 1 also presents the mean performance of *Ss* in the three conditions on various measures of recall. For each of these measures we performed a one-way analysis of variance, followed by three nonorthogonal comparisons. The first comparison, Sleep versus Awake, compared the pooled mean of the two sleep conditions (F and S) against the mean of the awake condition. The second comparison, F versus S, compared the first-half condition against the second-half condition. The third comparison, S versus A, compared the second-half mean with the awake mean.

TABLE 1

MEAN PERFORMANCE IN LEARNING AND RECALL

Response measure	Cond. F	Cond. S	Cond. A	MSwg	Significant comparisons[a]
Trials to criterion	8.10	7.00	7.30	17.30	—
Number correct on criterion trial	10.50	11.00	10.50	1.15	—
Absolute recall					
Test 1	8.30	6.40	4.60	4.48	1
Test 2	8.50	7.20	4.50	4.54	1, 3
Absolute loss					
Test 1	2.20	4.60	5.90	4.40	1, 2
Test 2	2.00	3.80	6.00	4.58	1, 3
Percent loss					
Test 1	20.90	41.94	56.34	377.17	1, 2
Test 2	19.14	34.50	56.97	383.29	1, 3
Number correct Unpaced test	9.90	8.30	5.70	5.08	1, 3
Number correct Matching test	12.10	11.00	8.80	6.98	1

Note.—MSwg is the mean square within groups from each analysis of variance. Comparison 1 = F + S versus A, 2 = F versus S, and 3 = S versus A.

[a] Significant at .05 for each of the dependent variables.

The column labeled "Significant comparisons" lists the number of the comparisons which were significant at the 0.5 level for each of the dependent variables.

Our main measures of retention were based on performance on the first of the two successive paced test trials. Three different measures were computed. The first, absolute recall, is the number of correct responses on this trial. The more crucial measures, however, are the two relative ones which we refer to as absolute loss and percent loss. Absolute loss is defined as the difference between the number correct on the criterion trial and number correct on the paced recall trial—it refers to the number of items lost over the retention interval. Percent loss is defined as the absolute loss divided by the number correct on the criterion trial times 100 and, of course, refers to percentage of items originally learned which was lost over the retention interval. Both these relative measures take into account slight differences among *S*s in degree of original learning and, as such, are the most appropriate and sensitive recall measures. Also, it is the first paced test which is of primary interest. The second paced test, the unpaced test, and the matching test represent successive retention measurements on the same *S*s, meaning that these results must be interpreted with caution. Table 1 presents the means for all the measures and lists the significant comparisons.

For absolute recall on the first test trial, the effect of condition was significant, $F(2, 27) = 7.65$, $p < .01$. Only comparison 1 reached significance, $F(1, 27) = 11.26$, $p < .01$, meaning that the pooled sleep conditions exhibited higher retention than the awake condition. The difference between the first- and second-half conditions approached significance ($p < .10$).

The picture is identical for both the absolute and percent loss scores on the first paced recall trial. The effect of conditions was significant, $F(2, 27) = 8.00$ and 8.42 for absolute and percent loss, respectively, $p < .01$. For both measures, Comparisons 1 and 2 were significant, but not Comparison 3. This means that retention was superior in the sleep groups compared to the awake group, and that within the two sleep conditions, retention was better in the first-half condition than in the second-half condition. The second-half condition, however, was not significantly different from the awake condition. Thus, these results exactly replicate the findings of Yaroush et al. (1971), while controlling for time of day.

On the second paced recall trial, the effect of condition was still significant, $F(2, 27)$

= 8.77 and 9.44 for absolute and percent loss, respectively, $p < .01$. However, now it is Comparisons 1 and 3 which are significant. This means that there was a substantial superiority of both sleep conditions over the awake condition, but no longer is there a significant difference between the two sleep conditions. A substantial numerical superiority of Cond. F over Cond. S still remains, however. The same pattern appeared on the unpaced test where the main effect of conditions was also significant, $F\ (2,\ 27) = 8.85$, $p < .01$. Each sleep condition showed significantly higher retention than the awake condition, but the two sleep conditions did not differ. Thus, the difference between the two sleep conditions was significant only on the first paced test when measured by absolute or percent loss.

On the matching test, the main effect of condition was significant, $F\ (2,\ 27) = 4.04$, $p < .05$. Only Comparison 1 was significant—the pooled sleep groups retained more than the awake group. This suggests that even when response availability is equated, sleep facilitates recall, implying that sleep can facilitate the retention of the association per se. The matching results must be interpreted with caution, however, since the matching test is the fourth test in a series of four successive retention tests given to all *S*s.

Sleep-Stage analyses.—Table 2 presents the results of the scoring of the physiological records. The data are expressed as the percentage of time spent in each of six stages (Wakefulness, I, II, III, IV, and REM). The first two columns represent the percentages of the various stages for the two sleep conditions during the 4-hr. retention interval. Column 3 shows the sleep percentages for the 4 hr. preceding the awake retention interval for the awake *S*s. Column 4 shows the sleep percentages for the 4 hr. preceding the retention interval in the second-half condition.

TABLE 2

PERCENT TIME IN VARIOUS SLEEP STAGES

Stage	Condition and time			
	F (2:50–6:50 a.m.)	S (2:50–6:50 a.m.)	A (10:30–2:30 a.m.)	S (10:30–2:30 a.m.)
Awake	6.13	14.04	7.76	11.87
I	6.38	6.77	5.39	6.16
II	41.64	45.39	35.68	41.42
III	4.64	3.98	6.08	5.34
IV	29.16	6.77	37.48	27.91
REM	11.23	22.57	6.69	6.71

It can be seen from Table 2 that the retention intervals in Cond. F and S were rather comparable on Stages 1, 2, and 3. The first-half *S*s did spend less time in wakefulness, probably due to the fact that they were slightly sleep deprived and could fall asleep faster than the second-half *S*s who already had had 4 hr. of sleep. This difference in wakefulness between Cond. F and S suggests that Cond. S remembered less because they slept less, something entirely consistent with the interference theory. Such an interference interpretation gains some support from a correlational analysis. We correlated amount of wakefulness with percent loss in the first-half condition, in the second-half condition, and in the pooled sample from both conditions. The pooled correlation, based on 20 *S*s, was significant, $r = .47$, $p < .05$; but the other two correlations, done on the separate conditions, were not ($r = .28$ and .35 in the first- and second-half conditions, respectively). Despite this, we are convinced that the F-S difference is not due to wakefulness differences because we have twice replicated the effect when there were no differences in wakefulness.[3] In neither of these replications were there significant correlations between amount of wakefulness and percent loss, and in one case the correlation was negative.

From Table 2 we can also see that the manipulations we employed to have the retention interval always occur between 2:50 a.m. and 6:50 a.m. did not destroy the basic sleep cycle which normally results in high amounts of Stage IV sleep and low amounts of REM sleep during the first half of the night and low Stage-IV and high REM during the second half. Cond. F had close to five times as much IV as Cond. S, while for REM sleep, Cond. S had almost twice as much as Cond. F.

---

[3] FOWLER, M. J., SULLIVAN, M. J., & EKSTRAND, B. R. Sleep and memory. Manuscript submitted for publication, 1972.

Discussion

The results of this experiment clearly indicate that the finding of superior retention over intervals filled with sleep as opposed to wakefulness is not an artifact of the confounding of sleep and wakefulness with time of day. For each of several measures of retention, including paced, unpaced, and matching tests, retention after sleep was significantly higher than after wakefulness, when the time of the retention interval was identical in both conditions. Furthermore, the results also strongly suggest that the difference in memory between the first and second halves of the night, first discovered by Yaroush et al. (1971), is also not due to the confounding with time of day. This is not to say that there are no time-of-day effects in memory, but only to show that such effects, if they exist, are not responsible for the better recall by sleep *S*s, and the better recall by first-half *S*s.

Contrary to the Yaroush et al. (1971) data, the present experiment strongly suggests that memory over the second-half of the night, while inferior to the first half, is superior to the awake condition. Yaroush et al. found no significant differences between the second-half and awake conditions on any of these recall measures. However, we did find a significant difference between these two conditions on four of our recall measures (see Table 1), particularly on the second of the two paced tests. Our other experiments (see Footnote 3) have also consistently shown a significant superiority of the second-half condition over the awake condition. We conclude that first-half memory is superior to second-half memory which is superior to memory over intervals of wakefulness.

These results, however, do not shed any new light on the difference between first and second halves of the night. The data on sleep stages strongly suggest the hypothesis that either Stage IV sleep facilitates memory or that Stage REM sleep inhibits it or both. Yaroush et al. (1971) suggested such a facilitation-disruption interpretation, hypothesizing that these stages affect a consolidation process. Another possibility is that REM sleep acts somewhat like interpolated experience of a general sort, having an interference-like effect and causing RI. Still a third possibility arises when we note that there is still a confounding present, both in this experiment and the Yaroush et al. (1971) experiment. The two conditions which result in the lowest recall, awake and second-half, are conditions in which the retention interval is preceded by at least 4 hr. of sleep. The first-half condition which results in highest recall is preceded by little or no sleep. It is possible that the prior sleep could in some way serve to depress recall; i.e., there is proactive sleep inhibition. Such a proactive effect could not explain the significant difference between the second-half condition and the awake condition which occurred in this experiment on some of the recall measures. This difference could be explained by saying that sleep also works by preventing IL, the interference-theory hypothesis.

It may also be noted here that the groups differed in the amount of sleep deprivation at the time of OL. The first-half *S*s were slightly sleep deprived at 2:30 a.m. whereas *S*s in the other two conditions had been sleeping since 10:30 p.m. Unless one wishes to take the position that prior sleep deprivation facilitates memory, this difference in deprivation will not account for the difference between the first- and second-half conditions.

We conclude from this experiment that there is a very reliable facilitating effect of sleep on memory which is not a time-of-day artifact. Furthermore, since we have replicated the Yaroush et al. (1971) results, showing superior recall over the first half of the night compared to the second half, we conclude that this is also a reliable effect. We also conclude that it will be a difficult chore to unravel the various possible explanations of these effects. Our best guess is that sleep can have effects on memory in more than one way, perhaps by preventing IL and reducing RI, perhaps by facilitating consolidation, or by interfering with consolidation. It is even possible that prior sleep interferes with memory or that prior deprivation of sleep facilitates it. Developing experimental designs which allow for the possible separation of these effects will not be easy.

References

Ekstrand, B. R. Effect of sleep on memory. *Journal of Experimental Psychology*, 1967, **75**, 64–72.

Jenkins, J. G., & Dallenbach, K. M. Obliviscence during sleep and waking. *American Journal of Psychology*, 1924, **35**, 605–612.

LUCE, G. G. *Biological rhythms in psychiatry and medicine.* (Tech. Rep. 2088) Chevy Chase, Md.: United States Department of Health, Education and Welfare, 1970.

RECHTSCHAFFEN, A., & KALES, A. (Eds.) *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.* (Tech. Rep. 204) Bethesda, Md.: United States Department of Health, Education, and Welfare, 1968.

VAN ORMER, E. B. Retention after intervals of sleep and of waking. *Archives of Psychology, New York*, 1932, No. 137.

YAROUSH, R., SULLIVAN, M. J., & EKSTRAND, B. R. Effect of sleep on memory. II: Differential effect of the first and second half of the night. *Journal of Experimental Psychology*, 1971, **88**, 361–366.

(March 13, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 328–333

PART–WHOLE LIST TRANSFER IN FREE RECALL:
A REAPPRAISAL

ROBERT E. HICKS[1] AND ROBERT K. YOUNG

University of Texas at Austin

Three part–whole transfer studies in free recall were conducted. In Exp. I, designed to be a replication of previous part–whole transfer experiments, a 9-item part list and then an 18-item whole list were learned. The control group learned unrelated lists of nouns, and the experimental group learned a whole list of nouns composed of the part-list items plus 9 new items. Experiment II was identical to Exp. I with the exception that adjectives were used instead of nouns. Experiment III was an exact replication of Exp. I and II combined except that the part lists were merely read aloud. Experiment I and the noun portion of Exp. III replicated previous research, while positive transfer was found in Exp. II and the adjective portion of Exp. III. Alternate explanations were considered.

The results of part–whole transfer studies in free-recall learning provide an interesting counterintuitive finding. In this paradigm, the experimental group learns a part list of N items and then learns a whole list of 2N items consisting of the part-list items plus N new items. The control group learns unrelated lists of N and 2N items. In contrast to what might be expected from more traditional theories of human learning (e.g., Asch & Ebenholtz, 1962; Underwood, 1964), negative transfer has been found (Tulving, 1966). Tulving's data can be easily explained by a theory of free recall (FR), which stresses the necessity for active organization of the list by the learner (Tulving, 1962).

Tulving (1966) reported a second counterintuitive finding which also supports organizational theory and which again is extremely difficult for a more traditional theory to explain. In this study, a "prior-acquaintance" group read aloud the items of a to-be-learned list six times, while a control group read an irrelevant list of words. Both groups then learned the test list by FR. In contrast to what might be expected, no difference was found between the two groups in the learning of the test list. Tulving's organizational theory can easily explain these data because it stresses that mere repetition alone is not an effective way to increase performance; rather, the material to be recalled must be organized into subjective groupings or *S* units.

The more traditional theories would predict positive transfer in both of the experiments reported by Tulving (1966) because of increased item availability and because individual items have been conditioned to the experimental context. The hypothesis of contextual associations has been applied primarily to paired-associate studies, e.g., Bilodeau and Schlosberg (1951) and McGovern (1964). Recent FR studies, however, have provided only limited support for the hypothesis, e.g., Abra and Belton (1969) and Strand (1970).

Just as paired-associate studies usually have used adjectives, FR studies have usually used nouns. The apparent reason for this tradition is that organizational processes in recall have long been of interest in research on FR (e.g., Bousfield, 1953), and nouns are most easily adapted to this type of research.

The purpose of the present paper is to examine transfer in part–whole learning when both adjectives and nouns are used so as to determine if the phenomenon observed with nouns can be found with another part of speech. In addition, the effect of pronunciation in a part–whole transfer paradigm is also studied. It is perhaps a measure of the impact of organizational theory upon this area of verbal learning that, to our knowledge, no replication of Tulving's (1966) study of the effects of prior acquaintance upon FR learning has ever been reported.

---

[1] Requests for reprints should be sent to Robert E. Hicks, Department of Psychology, University of Texas at Austin, Austin, Texas 78712.

Two part–whole FR learning studies, one using nouns and the other using adjectives, are reported. A third experiment, in which the part list was merely read aloud with no instructions to learn, employed both nouns and adjectives as the learning material.

EXPERIMENTS I AND II

Method

General design.—In both Exp. I and II, all *Ss* learned a 9-item list by FR and then learned an 18-item list by FR. The Control *Ss* learned two completely unrelated lists, while *Ss* of the Experimental groups learned a second (whole) list composed of all the items of the first (part) list plus an equal number of new items. Within each experiment all *Ss* learned the same whole list. Thus the design for both Exp. I and II can be summarized as a simple two-group design.

Specific procedure.—The lists used in Exp. I were composed of unrelated common nouns taken from lists prepared by Battig and Montague (1969), having a median Thorndike-Lorge G frequency of 25.00. Two groups of 34 *Ss* served in Exp. I. The lists used in Exp. II were composed of unrelated common adjectives taken from lists prepared by Melton and Safier (Hilgard, 1951). These lists had a median Thorndike-Lorge G frequency of 20.50. Two groups of 51 *Ss* served in Exp. II.

General procedure.—The verbal material was presented via a Kodak Carousel 650 projector with attached timer. All items in both lists and in both experiments were presented at a 1-sec. rate. Following each of the 10 study trials on the part list, there was a 30-sec. recall interval; and following each of the study trials on the whole list, there was a 60-sec. recall interval. Approximately 10 sec. elapsed between the end of a recall trial and the start of a study trial, and approximately 1 min. elapsed between the part and the whole lists.

The *Ss* were taken from introductory psychology classes and were tested in subgroups of five or six. During recall, *Ss* wrote their responses on a blank sheet of paper and covered their responses as they wrote them to prevent possible cueing from previously recalled items. At the end of each trial, the response sheet was placed face down on the table to expose a new blank sheet. There were five different orders of presentation for both the part and whole lists. Each subgroup was randomly assigned to a treatment group, and standard FR instructions were given. No *Ss* were dropped from the experiment.

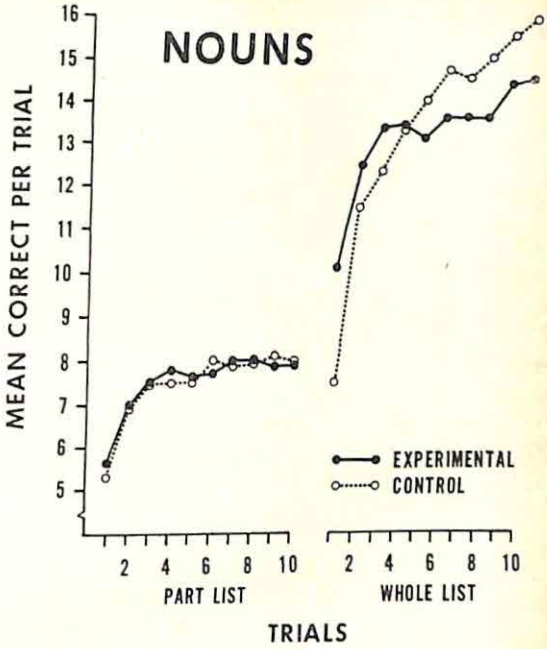


FIG. 1. Mean correct per trial during part and whole list learning in Exp. I.

Results

Exp. I.—The mean numbers of correct responses given during part-list learning are presented in Fig. 1. In the analysis only trials was significant, F (9, 594) = 50.19, $p < .001$. No other F was larger than 1.00, and the hypothesis of equivalent groups cannot be rejected.

The mean numbers of correct responses given during whole-list learning are also presented in Fig. 1. In this analysis both trials, F (9, 594) = 94.26, $p < .001$, and Trials × Groups, F (9, 594) = 9.88, $p < .001$, were significant. The former F indicates that whole-list performance improved over trials while the latter F replicates Tulving's (1966) finding of initial positive transfer followed by negative transfer.

Exp. II.—The mean numbers of correct responses given during part-list learning are presented in Fig. 2. In this analysis only trials was found to be significant, F (9, 900) = 71.09, $p < .001$. No other F was larger than 1.00, and again the hypothesis of equivalent groups cannot be rejected.

The mean numbers of correct responses

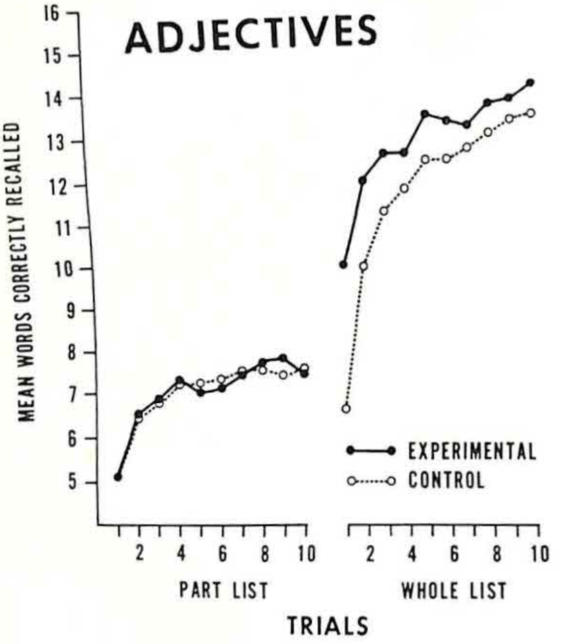


FIG. 2. Mean correct per trial during part- and whole-list learning in Exp. II.

given during whole-list learning are also presented in Fig. 2. In the analysis of these data, trials, F (9, 900) = 117.75, $p < .001$; groups, F (1, 100) = 9.15, $p < .01$; and Trials × Groups, F (9, 900) = 9.46, $p < .001$, were all significant. The significant F for trials simply indicates that whole-list performance improved over trials. The significant difference between groups indicates that positive transfer occurred during whole-list learning, with the mean number of adjectives recalled in the Experimental group being 12.80 per trial while the comparable mean in the Control group was 11.60. The Trials × Groups interaction indicates that the amount of positive transfer decreased over trials. It would appear that the most parsimonious explanation of this interaction would be in terms of a ceiling effect for the Experimental group but, interestingly enough, it can be taken as support for organizational theory. The reason for this is that the significantly more rapid improvement in performance for the control group may indicate the use of nonoptimal S units, formed during part-list learning by the Experimental group.

EXPERIMENT III

Method.—Each of 68 *S*s read aloud a part list and then learned a whole list. Half of the *S*s read a part list composed of items subsequently learned in the whole list while the other half of the *S*s read an irrelevant part list. Within each of these groups, half of the *S*s read and learned lists composed of nouns while the other half read and learned lists composed of adjectives. Thus the design can be summarized as a 2 × 2 factorial with relevance of the part list (experimental vs. control) and grammatical class (adjective vs. noun) being combined factorially ($n = 17$ per cell).

The only difference between Exp. III and the comparable groups of Exp. I and II was that the part lists were merely read aloud as they were continuously presented. Again, the part and whole lists were presented for 10 trials. The noun lists were the same as in Exp. I and the adjective lists were the same as in Exp. II. The *S*s in Exp. III were selected in the same manner from the same population and assigned to treatments the same way as those in Exp. I and II.

The *S*s were again run in subgroups of five and six at a time. Before presentation of the part list, *S*s were told that a series of words would appear on the screen and that they were to read each word aloud as it appeared. The order of presentation was exactly the same as in Exp. I and II. The words were presented continuously at a 1-sec. rate. The *S*s tended to pronounce the words in unison. Following presentation of the part list, standard FR instructions were given and whole-list learning began immediately thereafter.

Results.—The mean numbers of correct responses in the four groups of Exp. III are presented in Fig. 3. It can be seen that several differences were obtained. The noun lists were easier to learn than the adjective lists, F (1, 64) = 33.14, $p < .001$. Similarly, positive transfer occurred, with the experimental lists being easier to learn than the control, F (1, 64) = 10.18, $p < .01$. These two variables interacted, and positive transfer was found with the adjectives but not with the nouns, F (1, 64) = 9.44, $p < .001$. Performance increased over trials, F (9, 567) = 64.17, $p < .001$, and amount of transfer changed over trials such that the Trials × Relevance (of the Part List) × Grammatical Class interaction was highly significant, F (9, 576) = 5.72, $p < .001$. Figure 3 can be used to better understand the meaning of this interaction. It can be seen there that positive transfer occurred throughout the course of learning for the adjective groups, while positive transfer occurred only in the early trials for the noun groups. Subsequently, this positive transfer in the noun groups disappeared. Thus the results of Exp. III replicate those found in Exp. I and II. Finally, the Trial × Grammatical Class interaction was significant, with nouns being learned somewhat more rapidly than adjectives, $F\ (9, 576) = 2.34, p < .05$.

It should be noted that the noun portion of Exp. III replicates the results found by Tulving (1966) in his experiment in which *S*s merely read words aloud. That is, an analysis of the noun data alone in Exp. III failed to find a significant difference between the Experimental and Control groups. Similarly, the Trials × Revelance (i.e., Experimental vs. Control) interaction also failed to attain significance.

DISCUSSION

The results of Exp. I replicate Tulving's (1966) finding of negative transfer and are in agreement with Bower and Lesgold (1969), Novinski (1969), and Roberts (1969). All the studies listed above used nouns, as have most other FR studies, and all of them found initial positive transfer followed by negative transfer in a part–whole paradigm. These results support Tulving's (1966) organizational theory of FR learning.

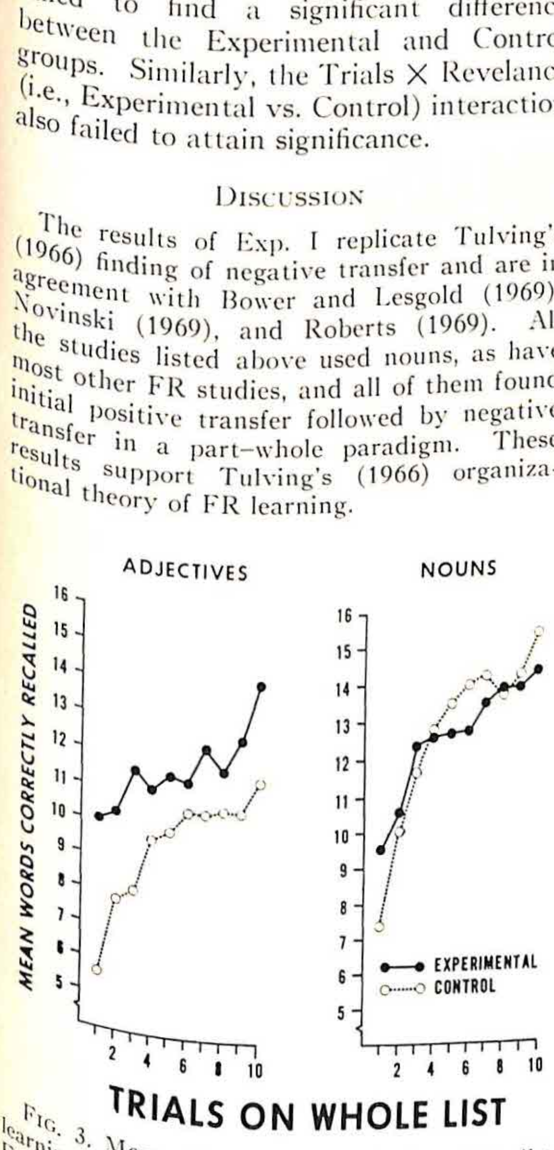


FIG. 3. Mean correct per trial during whole list learning for the noun and adjective groups in Exp. III.

Similarly, the part of Exp. III employing nouns also replicates Tulving's (1966) finding that "mechanical repetition" i.e., reading the items aloud with no instructions to learn, has no effect on subsequent learning and recall. According to his organizational theory, only when repetition leads to the formation of higher order *S* units is it effective in allowing *S* to recall more items than on the first trial. When *S*s are not instructed to memorize, then such higher order *S* units will not be formed, and the learning of an FR list will be no more rapid than that of a control FR list.

Experiment II and the part of Exp. III employing adjectives fail to unequivocally support the organizational hypothesis. It should be recalled that Exp. II was identical to Exp. I except that adjectives instead of nouns were used, but the typical results found in other part–whole studies—positive transfer followed by negative transfer—was not found. Instead, consistent positive transfer was obtained throughout whole-list learning. These data fit the more traditional theories better than they fit organizational theory. The traditional theories emphasize response availability and contextual association, and these theories would predict that positive transfer should occur during whole-list learning since half of the experimental items have already had their availability increased and been conditioned to the experimental context. Similarly, in Exp. III, when the items are merely repeated again and again without instructions to memorize, both availability and contextual associations should be increased yielding positive transfer.

While organizational theory can predict the positive transfer obtained in Exp. II simply by arguing that optimal *S* units were formed in part-list learning (cf. Slamecka, Moore, & Carey, 1972), it appears to be somewhat more difficult for organizational theory to explain the data of Exp. III. In this experiment, *S*s merely read the part-list words as they were presented. A theory which stresses increased availability of responses would predict positive transfer with both nouns and adjectives, while organizational theory would predict zero transfer in both cases. Positive transfer was obtained with adjectives and zero transfer was obtained with nouns, and it can be seen that the data which cannot be explained by organizational theory are easily accounted for by availability theory and vice versa. Thus neither theory can give an adequate account of the results obtained in Exp. III.

Adjective lists are more difficult to learn than noun lists. This may simply be due to the fact that nouns tend to have a higher degree of imagery than adjectives. If the results obtained in the present series of experiments are attributable to this difference, then a comparison of high- and low-imagery nouns in a part–whole transfer paradigm should yield differential transfer similar to the present results. In addition, the finding of differential transfer would suggest that organizational factors are most important when the relative difficulty of the learning material is low (i.e., when imagery is high) and that associative factors are most important when the relative difficulty of the learning material is high.

The differences obtained in the present series of studies may not be related so much to task difficulty as to word class. Nouns and adjectives have been used by *S* in quite different ways prior to the experiment, and the manner in which an item has been used by *S* may be related to the pattern of transfer obtained. To the *S*, nouns may represent a collection of items, while adjectives may represent a collection of attributes. Since it is not unusual for *S* to be familiar with concepts with overlapping sets of attributes, e.g., a *big, round orange* and a *big, round, ripe, wormy apple,* it may be that transfer of adjectives already available is easier than the transfer of nouns. If the results of the present experiment are attributable to differences in word class, then a comparison of adjectives and nouns with imagery held constant should yield results comparable to the differential transfer obtained in the present series of experiments.

Another explanation can be based on the results of Postman (1971), who found that associations, formed among items of an FR list, are strong enough to produce negative transfer in subsequent learning. Assume that preexperimental internoun associations are relatively strong, while interadjective associations are relatively weak. Under these conditions, an adjective list would be more difficult to learn than a noun list. With the addition of new items to the adjective list, as in the part–whole transfer paradigm, the associations formed during part-list learning would facilitate subsequent whole-list learning. However, with the addition of new items to the noun list, relatively strong preexperimental associations between part-list items and new items would interfere with the associations formed and/or strengthened during part-list learning. That is, BAT and BALL might be easily associated in part-list learning. With the addition of WING and DANCE in whole-list learning, the BAT–WING and DANCE–BALL associations would tend to interfere with the BAT–BALL association, resulting in retarded whole-list learning.

The results of the present series of experiments along with those of Novinski (1972) and Slamecka et al. (1972) suggest that organizational theory has severe limitations on its application. The usual part–whole transfer effect can be modified considerably through the use of different instructions or different verbal materials. While organizational theory can explain these results, it does so at the risk of becoming overly general. On the other hand, while the more traditional theories emphasizing response availability and contextual associations have continual difficulty in explaining the typical part–whole transfer result, they can readily handle the results of the adjective portions of the present series of experiments.

REFERENCES

Abra, J. C., & Belton, P. S. Unlearning in free learning as a function of context change and retention interval. *Psychological Reports,* 1969, **24,** 943–946.

Asch, S. E., & Ebenholtz, S. M. The process of free recall: Evidence for non-associative factors in acquisition and retention. *Journal of Psychology,* 1962, **54,** 3–31.

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology,* 1969, **80** (3, Pt. 2).

Bilodeau, I. McD., & Schlosberg, H. Similarity in stimulating conditions as a variable in retroactive inhibition. *Journal of Experimental Psychology,* 1951, **41,** 199–204.

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology,* 1953, **49,** 229–240.

Bower, G. H., & Lesgold, A. M. Organization as a determinant of part-to-whole transfer in free recall. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8,** 501–506.

Hilgard, E. R. Method and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

McGovern, J. B. Extinction of associations in four transfer paradigms. *Psychological Monographs,* 1964, **78** (16, Whole No. 593).

Novinski, L. Part-whole and whole-part free recall learning. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8,** 152–154.

NOVINSKI, L. S. A reexamination of the part/whole effect in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 228–233.

POSTMAN, L. Organization and interference. *Psychological Review*, 1971, **78**, 290–302.

ROBERTS, W. A. The priority of recall of new items in transfer from part-list to whole-list learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 645–652.

SLAMECKA, N. J., MOORE, T., & CAREY, S. Part-to-whole transfer and its relation to organization theory. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 73–82.

STRAND, B. Z. Change of context and retroactive inhibition. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 202–206.

TULVING, E. Subjective organization in free recall of "unrelated" words. *Psychological Review*, 1962, **69**, 344–354.

TULVING, E. Subjective organization and effects of repetition in multi-trial free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 193–197.

UNDERWOOD, B. J. The representativeness of rote verbal learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964.

(Received March 15, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 334–337

PREDICTION OUTCOME AND CHOICE REACTION TIME: A MEMORY-DEPENDENT RELATIONSHIP[1]

E. SCOTT GELLER[2] AND CHARLES P. WHITMAN

Virginia Polytechnic Institute and State University

In two-choice reaction time (RT) situations, the interval between the prediction and the presentation of successive stimuli was either 3, 7, or 11 sec. The 20 *S*s of one group (Group Stim) pressed a left-hand trigger following one stimulus and a right-hand trigger following the alternative stimulus; *S*s of another group (Group Pred) identified the prediction outcome on each of the 200 trials by making one response if the prediction was correct and the other response following an incorrect prediction. The influence of the prediction–stimulus interval implicated both expectancy and memory factors in choice RT. That is, RTs were faster following correct than incorrect predictions, and this effect of prediction outcome was greatest following the shortest interval. The reliable effects of prediction outcome on RT were more prominent for Group Pred than for Group Stim.

When required to predict the sequential occurrence of stimulus alternatives in a choice reaction time (RT) paradigm, *S*s react markedly faster to correctly predicted than to incorrectly predicted events (e.g., Geller & Pitz, 1970; Geller, Whitman, & Farris, 1972; Keele, 1969). The present experiment was designed to determine if the temporal interval occurring between the prediction and the presentation of a stimulus would influence the effects of the prediction outcome on reaction latency. The three prediction–stimulus intervals used (3, 7, and 11 sec.) were within a range in which short-term memory (STM) for consonant trigrams was shown to decay (Peterson & Peterson, 1959). Hence, an STM model might hypothesize that the prediction outcome effect would be most apparent at the shortest interval (i.e., 3 sec.), since at this interval the memory trace of a verbalized expectancy would be strongest. Following the longest interval (i.e., 11 sec.) the effect of prediction outcome on choice RT should be least prominent, since the verbal STM for the stimulus prediction would have faded or decayed to the greatest extent. Such effects of the time interval between the prediction and the presentation of a stimulus would implicate both expectancy and memory factors in the choice RT situation.

Recent choice RT investigators have varied the amount of time between successive stimulus presentations in order to study the influence of STM on the "repetition effect" (i.e., faster reactions to the second of two consecutive identical rather than dissimilar stimuli). In both a four-choice and a two-choice RT task with one-to-one S–R mapping, Smith (1968) found the repetition effect to be a decreasing function of the interstimulus intervals: 2, 6, and 10 sec. Smith attributed her findings to a direct relationship between the length of the interval and the degree of fading of the stimulus trace. However, it should be noted that the results of Schvaneveldt and Chase (1969) and Keele (1969) did not support an STM interpretation of sequential effects in similar choice RT tasks, since the repetition effect did *not* decrease over the range of intervals studied (i.e., 100 msec. to 8.5 sec.).

The present experiment studied the prediction outcome effect as a function of the prediction–stimulus interval in two different RT situations: a standard two-choice RT task with one-to-one S–R mapping

---

[1] The authors thank David A. Williams for his assistance in collecting the data and Western Electric Company, Inc., for its donation of electronic supplies.

[2] Requests for reprints should be sent to E. Scott Geller, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

and an RT paradigm which required *S*s to identify the outcome of their predictions by making one response following correctly predicted stimuli and another response following incorrect predictions. Thus, the effects of prediction outcome and prediction–stimulus interval were determined in two task situations which perhaps involve dissimilar cognitive processes or a differential biasing of the same processes.

Method

Subjects.—The 40 *S*s were volunteers from the introductory psychology classes at Virginia Polytechnic Institute and State University. The *S*s received optional research credit for their participation in individual 1-hr. experimental sessions. Ten males and 10 females served in each of two experimental conditions.

Design and procedure.—During individual 1-hr. sessions, *S*s received 198 stimulus presentations, each being the symbol ⊔ or ⊓ (labeled UP and DOWN, respectively). The sequence of stimuli was derived by filtering a uniform random number generator on an IBM 360 computer so that the two stimulus alternatives were equiprobable (i.e., 99 UPs and 99 DOWNs).

Prior to each stimulus presentation, *S*s verbally predicted which alternative they expected to occur (UP or DOWN) and then with their left forefinger depressed one of two levers to indicate their prediction. The two prediction levers were located on a panel to the left of *S* and were separated on a vertical axis by 7.7 cm. To designate a prediction of UP, *S*s pressed the uppermost lever; the lower lever was pressed following verbal predictions of DOWN. Depressing a prediction lever initiated a time interval of 3, 7, or 11 sec., after which the stimulus was presented. Programmed by a punched paper tape, each of the three time intervals preceded an UP and a DOWN presentation on 33 trials. The distribution of intervals was the same for each *S* and was ascertained by the appropriate filtering of a random number generator on an IBM 360 computer.

Following a stimulus presentation *S*s made a choice reaction by pressing one of two microswitches, one mounted in a handle for *S*s' left hand and the other mounted in a right-hand trigger. Twenty *S*s in one group (Group Stim) identified the stimulus presented by making one response to ⊔ and the other to ⊓. For this condition, 10 *S*s pressed the left trigger following ⊔ and the right trigger following ⊓, while 10 *S*s used the reversed S–R mapping relationship. The 20 *S*s of a second group (Group Pred) identified the prediction outcome determined by each stimulus presentation. Ten *S*s used the left-hand trigger to identify a correct prediction and the right-hand trigger to identify an incorrect prediction; conversely, 10 *S*s pressed the left trigger following incorrect predictions and the right trigger following correct predictions.

The reaction latencies (between stimulus presentations and identification responses) were recorded to the nearest millisecond by a Hunter Klockounter, Model 120A. Whenever inaccurate identifications occurred, *E* told *S* of the error. After recording a reaction latency, *E* pressed a button to present a horizontal line (1.3 cm.) in the center of the stimulus readout (Allied No. 448MS14A), which served as a cue for *S* to make the next prediction. The horizontal line was not a section of either stimulus alternative (⊔ or ⊓).

The *S*s were encouraged to concentrate on their stimulus predictions and were told to take as long as they wanted to make each prediction. Also, *S*s were instructed to make the identification responses as quickly as possible while minimizing the number of errors. The *E* specified that the sequence of stimuli would be determined by a punched paper tape. No mention was made of the varying time intervals between prediction and stimulus, nor of the relative frequencies of the stimulus alternatives. The *S*s were given 10 practice trials, during which questions concerning their task were answered.

Results

The latencies for anticipations and inaccurate identifications were disregarded in the analysis and represented less than 5% of any *S*'s data. For each *S*, the reaction latencies were classified according to the three prediction–stimulus intervals and the prediction outcome determined by *S*'s prediction and the presented stimulus (correct or incorrect). Thus, each choice RT was placed into one of six categories indicated by the factorial of 2 (Prediction Outcomes) × 3 (Intervals). The group means depicted in Fig. 1 were computed by calculating the category means for each *S* and then averaging across *S*s' means for each group.

As illustrated in Fig. 1, the choice reactions of Group Stim were markedly faster than those of Group Pred, and for both groups choice RT was notably faster to a correctly predicted stimulus than to an incorrectly predicted stimulus. However, the latency difference due to prediction outcome appears to be greater for Group Pred than Group Stim. The analysis of variance, a factorial of 2 (Groups) × 2 (Prediction Outcomes) × 3 (Intervals), demonstrated significant main effects of group, F (1, 38) = 26.94, and prediction

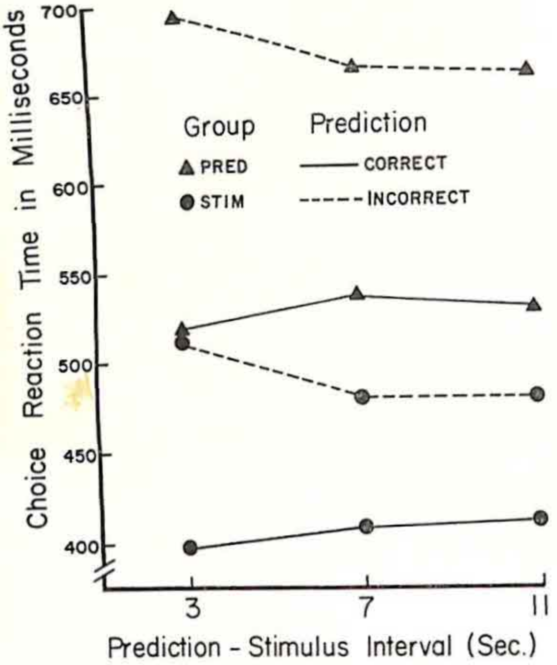


FIG. 1. Choice reaction time as a function of the identification response (Group Stim vs. Group Pred), the interval between prediction and stimulus, and the prediction outcome.

outcome, F (1, 38) = 163.97, both ps < .001. Only two significant interactions were obtained: Group × Prediction Outcome, F (1, 38) = 12.99, and Prediction Outcome × Interval, F (2, 76) = 14.88, both ps < .001.

As depicted in Fig. 1, the Prediction Outcome × Interval interaction was due to the greatest effect of prediction outcome on choice RT following the shortest prediction–stimulus interval (i.e., 3 sec.). That is, given a correct prediction, choice RT was faster when the prediction–stimulus interval was 3 sec. than when it was 7 or 11 sec., but for incorrect prediction outcomes choice RT was slowest when the prediction–stimulus interval was 3 sec. The separate analysis for correct prediction, 2 (Groups) × 3 (Intervals), showed significant main effects of group, F (1, 38) = 24.65, $p < .001$, and interval, F (2, 76) = 4.26, $p < .025$. The Group × Interval interaction was not significant, F (2, 76) = 2.56, $.10 > p > .05$. Likewise, the analysis for incorrect predictions indicated significant main effects of group, F (1, 38) = 26.11, $p < .001$, and interval, F (2, 76) = 6.43, $p < .005$, with a nonsignificant interaction term, F (2, 76) = .024.

DISCUSSION

The present findings not only support the importance of verbal predictions as determinants of choice RT, but also indicate a significant interaction between expectancy and memory factors. That is, whether the stimulus or the prediction outcome was identified, the influence of a verbal expectancy on a subsequent RT was most evident when the temporal interval between the prediction and the presentation of a stimulus was the shortest (i.e., 3 sec.). Indeed, the Interval × Prediction outcome interaction is supportive of an STM model similar to that used by Smith (1968) to explain her observation that the repetition effect on choice RT was reliably greater for the shortest interstimulus interval (i.e., 2 sec.). In particular, when a stimulus is presented for identification, the memory trace of a verbal expectancy is strongest when the time interval occurring between the prediction and the stimulus has been relatively short. Hence, it might be assumed that the degree of facilitation due to correct predictions and inhibition due to incorrect predictions are directly related to the strength of the verbal memory trace. Since the prediction outcome effect was generally of the same magnitude for the two longer intervals (i.e., 7 and 11 sec.), the maximum degree of memory fading as reflected in choice latencies apparently occurred between 3 and 7 sec.

The observation that the prediction outcome had a greater effect on reaction latency for Group Pred than for Group Stim supports a notion that prediction outcome effects are dependent on task complexity. Keele (1969) and Whitman and Geller (1971) varied S–R compatibility between *S*s and demonstrated a significantly greater effect of prediction outcome under the incompatible condition. Likewise, in the present study the latency differences due to prediction outcome were reliably larger for the RT task which resulted in markedly longer identification times (i.e., Group Pred). Incompatible tasks are assumed to emphasize response factors to a greater extent than compatible tasks (e.g., Geller et al., 1972; Whitman & Geller, 1971). Perhaps, response selection is also a more important process when *S*s identify prediction outcomes than when they identify stimuli.

REFERENCES

GELLER, E. S., & PITZ, G. F. Effects of prediction, probability and run length on choice reaction speed. *Journal of Experimental Psychology*, 1970, **84**, 361–367.

GELLER, E. S., WHITMAN, C. P., & FARRIS, J. C. Probability discrimination indicated by predictions and reaction speed: Effects of S–R compatibility. *Journal of Experimental Psychology*, 1972, **93**, 404–409.

KEELE, S. W. Repetition effect: A memory-dependent process. *Journal of Experimental Psychology*, 1969, **80**, 243–248.

PETERSON, L. R., & PETERSON, M. S. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

SCHVANEVELDT, K. W., & CHASE, W. G. Sequential effects in choice reaction time. *Journal of Experimental Psychology*, 1969, **80**, 1–8.

SMITH, M. C. Repetition effect and short-term memory. *Journal of Experimental Psychology*, 1968, **77**, 435–439.

WHITMAN, C. P., & GELLER, E. S. Prediction outcome, S–R compatibility, and choice reaction time. *Journal of Experimental Psychology*, 1971, **91**, 299–304.

(Received March 20, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 338–344

EFFECT OF STIMULUS RATE, MATERIAL, AND STORAGE INSTRUCTIONS ON RECALL OF BISENSORY ITEMS:

STORAGE OR RETRIEVAL EFFECTS?[1]

PAMELA C. FREUNDL
University of California, Los Angeles

GERALD M. SENF[2]
University of Iowa

An audiovisual analogue of dichotic listening, consisting of four pairs of simultaneous, discrepant items, examined the effects of material (letters vs. digits), interstimulus interval (.75 vs. 2 sec.), and storage instructions on bisensory recall and recognition memory. Previous recall studies showing retrieval in channel sets to characterize fast stimulus presentation rates and pair (actual) order retrieval to characterize slow rates have not demonstrated conclusively that these retrieval order differences reflect underlying storage differences. Comparison of recognition versus recall data supports the conclusion that storage of rapidly presented items proceeds in channel (modality) sets, while storage can be accomplished either in pair or channel groupings at the slower stimulus rate. Rapid stimulus rates likely restrict the manner of storage for this type of material through an inherent limitation in *S*'s ability to shift attention from one stimulus modality to the other. Slow presentation rates allow sufficient time to switch attention and, hence, variously to order the items in memory.

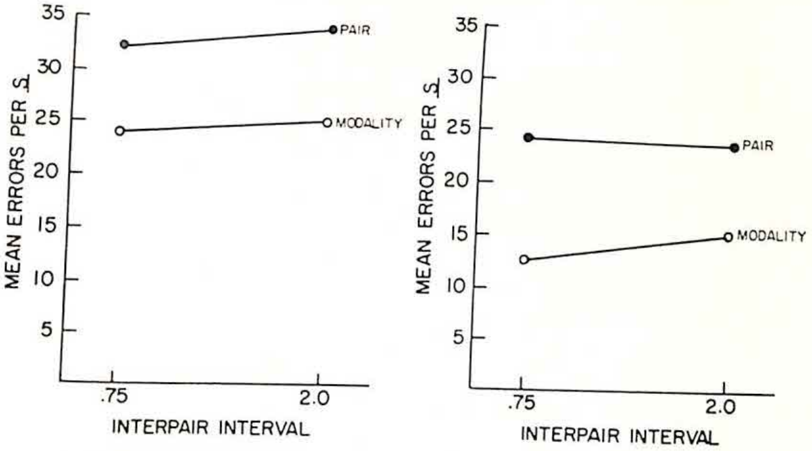
In two previous series of studies (Madsen, Rollins, & Senf, 1970; Senf, 1969; Senf & Freundl, 1971; Senf & Rollins, 1971), we have experimented with an audiovisual analogue of dichotic listening to test certain aspects of Broadbent's (1957) model of human memory and attention and to ascertain whether functions assessed by this bisensory analogue discriminated adequate readers from those afflicted with specific learning disabilities. The bisensory technique administers simultaneous, discrepant audiovisual pairs of items to *S* for immediate recall, usually three such pairs being administered on a trial with stimulus rate varying from less than one to as many as three pairs per second. In general, this audiovisual analogue has generated results (Madsen et al., 1970) comparable to the dichotic procedure used by Broadbent (1954) and more recently by others (e.g., Bryden, 1962; Treisman, 1969).

An initial experiment in the series reported here sought mainly to assess whether slow readers of college age evidenced deficits functionally similar to reading-retarded children. Though the samples of adequate and deficient readers did not differ, the working of the experimental variables was provocative. The *S*s were administered bisensorial stimuli of two rates (.75 sec. or 2 sec. between successive pairs) and of two material types (letters or digits) for recall in one of two recall orders (pair order or modality order). All *S*s received all conditions. The interpair interval and recall order variables were counterbalanced while letter and digit trials were randomly interspersed. The possible sequence effects allowed by this design, while not evaluated, warrant caution in accepting the results of this preliminary experiment. Each trial consisted of four pairs of bisensory items, either all digits or all letters, making eight unique items per trial. Pair recall condition dictated that *S* report the numbers starting with the initial visual stimulus followed by the initial auditory stimulus, followed by the second visual stimulus and so on. Modality recall order dictated that

---

[1] This research was supported by Biomedical Sciences Support Grant FR-07035 from the General Research Support Branch, Division of Research Resources, Bureau of Health Professions Education and Manpower Training, National Institutes of Health, to the second author while at the University of Iowa.

[2] Requests for reprints should be sent to Gerald M. Senf, who is now at the Department of Psychology, University of Illinois at Chicago Circle, Chicago, Illinois 60680.



A. LETTERS

B. DIGITS

FIG. 1. Mean errors per *S* as a function of stimulus material, interpair interval, and directed recall conditions for the preliminary experiment. (Each point is the mean of 48 scores; maximum score is 64.)

S first report the four visual stimuli followed by the four auditory stimuli. An error was scored for each stimulus item absent from the response or not recalled in the numerical position dictated by the recall order.

Analysis of variance found the pair recall order more difficult than the modality order, a typical finding (Madsen et al., 1970, Exp. VII; Senf, 1969, Exp. II & III; Senf & Feshbach, 1970, Exp. II; Senf & Rollins, 1971). More notable results were that letter recall was inferior to digit recall, F (1, 46) = 333.1, $p < .001$ (letter mean = 29.03, SD = 10.37; digit mean = 18.78, SD = 11.25), and that material interacted with recall condition and interpair interval. This second effect is notable for its weakness, F (1, 46) = 4.43, $p < .05$. As seen in Fig. 1, the triple interaction is barely discernable. Completely absent was the usual interaction between recall order and interpair interval, the original finding by Broadbent (1954) in the dichotic listening paradigm, whereby recall of rapidly presented materials is more accurate in channel (here modality) order while recall of slowly presented material is superior in pair order (Madsen et al., 1970, Exp. VII; Senf & Rollins, 1971). Thus, the usual effect of stimulus rate found using six bisensory items was markedly attenuated using eight items. Whereas Broadbent (1957) has interpreted the Stimulus Rate × Retrieval Method interaction as reflecting the organism's natural tendency to store rapidly presented materials in channel sets and slowly presented materials in their order of arrival, failure to find this effect (together with the weak third-order effect with material) raises the issue of whether increasing the number of items has actually changed the method of item storage. Alternatively, the manner of storage may have proceeded as in the six-item case but the difficulty of retrieving eight items may have attenuated the usual interaction. Likewise, the fact that letters were found to be more difficult to recall than digits may stem from greater difficulty in retrieving the more information-laden letters than from greater difficulty in storing them.

The following experiment assessed whether the attenuation of the Stimulus Rate × Recall Order interaction and the greater difficulty in recalling letters result more from storage or retrieval processes. We assumed that a partial-probe recognition procedure markedly reduces the role of retrieval processes (Kintsch, 1968; Murdock, 1968). If the attenuation of the rate effect results from retrieval processes, a strong Pair versus Modality × Stimulus Rate interaction should result. If letters are harder to recall than digits by virtue of retrieval processes, the recognition procedure should find them as easily recognized as digits.

We also addressed the additional issue of whether *S* can control his method of bisensory item storage independent of the stimulus presentation rate. Broadbent (1957) has argued that the rate at which items arrive at two sensory channels determines the manner of item storage and, hence, the accuracy of various recall orders. Items arriving at rapid rates, e.g., interpair intervals of approximately .5–1 sec., are easily recalled in channel sets but not as ordered pairs (actual order of arrival). Conversely, items arriving in pairs spaced further apart in time, e.g., interpair intervals of 2 sec., are more easily recalled as pairs than as channel sets. The authors' previous research has found these predictions, based primarily in the study of dichotic, binaural channels (Broadbent, 1954; Bryden, 1962) applicable to the bisensory situation (Madsen et al., 1970, Exp. VII; Senf & Rollins, 1971). Further, recall order preference data (Madsen et al., 1970, Exp. I-IV), as opposed to recall accuracy, are also in line with Broadbent's (1957) mechanical model. However, other data (Senf, Rollins, & Madsen, 1967) have shown that instructional sets administered prior to stimulus presentation can override the effect of stimulus rate on recall order. Thus, *S*s induced to view six bisensory items as audiovisual pairs recalled the items in pair (actual) order irrespective of the interpair interval; conversely, *S*s induced to view the items as modality sets recalled the items in two channel groupings, again irrespective of the interpair interval. The authors (Senf et al., 1967) have interpreted these results to mean that *S* can actually store items in accord with either organizational scheme and that the primary effect Broadbent has assigned to stimulus rate is inappropriate, though other data have found that the effectiveness of the set manipulation was limited by task difficulty (Madsen et al., 1970, Exp. VI) making this interpretation more tenuous.

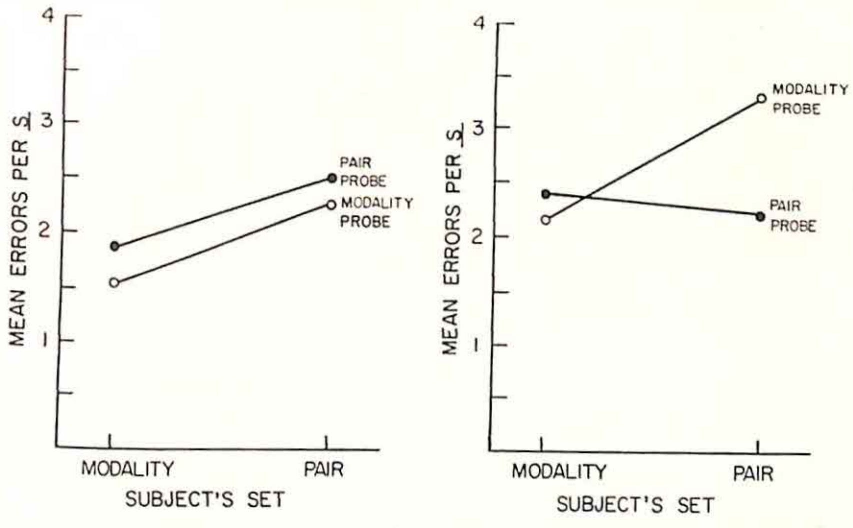
The experiment reported here used eight bisensorial stimuli but required that *S* only *recognize two* of the stimulus items instead of *recalling all eight* items. On each trial, a two-item set was probed by presenting it with two foil item sets. This procedure can be thought to bypass or at least minimize retrieval processes. By probing items, either adjacent within a modality or audiovisual pairs, one can assess whether *S* can control his storage independent of the stimulus rate. For example, probing an audiovisual pair from eight rapidly presented items, after directing *S* to store the items as pairs, should result in increased accuracy relative to probing adjacent items within a modality unless *S* is unable to modify the supposedly "natural" modality storage appropriate to the fast stimulus rate (Broadbent, 1957). Similarly, probing adjacent items within a modality with the stimuli presented slowly, after directing *S* to store the items as modality sets, should result in increased accuracy relative to probing an audiovisual pair unless *S* is unable to modify the supposedly "natural" pair storage appropriate to the slow stimulus rate (Broadbent, 1957).

EXPERIMENT I

Method

Subjects.—The 60 *S*s were students enrolled in elementary psychology classes at the University of Iowa. An equal number of males and females were randomly assigned to treatments, but sex was not analyzed as a variable.

Apparatus and stimulus materials.—Each trial contained four pairs of simultaneous, discrepant auditory and visual digits or letters arranged in audiovisual pairs. Half the trials were composed of digits and half of letters. The items for a particular trial were selected without replacement so that a given stimulus occurred only once per trial. The visual items were printed in .635-cm. black, block type on Bell & Howell "Language Master" cards. The auditory items were recorded on audiotape. These materials were simultaneously presented by means of a Language Master (#711B). Additional materials included probe-recognition cards presented immediately after the stimulus presentation. Each card, one per trial, contained three grids composed of two rows of four immediately adjacent boxes. The boxes represented spatially the eight stimuli, the top row corresponding to the four visual items, the bottom row to the four auditory items. Two boxes of one grid contained items accurately representing the items of the stimulus. The corresponding boxes in the other grids contained inaccurate items and, hence, were foils. The remaining six boxes in each grid were left blank. The *S*'s task was simply to



A. .75-SEC. INTERPAIR INTERVAL B. 2.0-SEC. INTERPAIR INTERVAL

FIG. 2. Mean errors per *S* as a function of interpair interval, instructional set, and probe type for Exp. I. (Each point is the mean of 15 scores; maximum score is 8.0.)

indicate which of the three grids contained the items which had appeared in those grid locations in the stimulus.

The two foils on each probe-recognition card had the following form: One item was correct while the other was selected from one of the other six items presented in the stimulus. Each foil contained a different correct item. Selection of the incorrect item composed an additional variable; the incorrect item was selected either from the three remaining items of the same modality as the correct item it replaced (nonshift lure) or from the three items of the opposite modality (shift lure). Both foils on any given trial were of the same lure type.

The two-item probe was either the modality type or the pair type. A modality probe consisted of two adjacent items within the same modality; a pair probe consisted of adjoining (simultaneous) items from different modalities. Each probe type appeared in four locations. Modality probe locations were the first two visual items, the second two visual items, the first two auditory items, and the second two auditory items. Pair probe locations were the four audiovisual pairs. Thus, all *Ss* received factorial combinations of Probe Types (pair vs. modality) × Probe Locations × Lure Type (shift vs. nonshift) × Stimulus Materials (letters vs. digits). Letters and digits were administered in counterbalanced blocks of 40 trials resulting in a sequence-of-materials variable that was not analyzed. Probe location also was not analyzed.

Independent groups received trials with stimuli presented at either a .75-sec. interpair interval or a 2-sec. interpair interval and under one of two instructional conditions described immediately below.

Procedure.—An *S*'s set was experimentally induced by means of instructions requesting that *S* attempt to remember the items either as modality sets, all the visual items together and all the auditory items together, or to remember the pairs as they occur, a visual and an auditory item together. The *Ss* were encouraged to encode the items in this manner even if it did not appear to be the most efficient strategy. To encourage *S* to use the designated encoding strategy, he was administered a disproportionate number of probe-type trials consistent with the strategy he was to adopt. Thus, for example, *Ss* instructed to store items in pairs received 50% more pair probe than modality probe trials. These additional trials were deleted from the data analysis.

Immediately after the stimulus presentation, *S* viewed the probe-recognition card and orally indicated whether the top, middle, or bottom grid was the appropriate representation of the preceding stimulus.

Results

The means reported are average errors per *S*, maximum = 8.0. Question type interacted with set, F (1, 56) = 6.27, $p < .05$, modality set improving performance on modality questions ($\bar{X} = 1.84$; $SD = 1.9$) compared to the Pair Set group's performance on modality questions ($\bar{X} = 2.80$; $SD = 1.58$) as predicted. The Pair Set group, however, produced slightly more errors on pair questions ($\bar{X} = 2.36$; $SD = .92$) than the Modality Set group ($\bar{X} = 2.13$; $SD = 1.67$). Question type also interacted significantly with interpair interval, F (1, 56) = 5.99, $p < .05$, the long interval producing more errors for modality questions ($\bar{X} = 2.73$; $SD = 1.44$) than pair questions ($\bar{X} = 2.30$; $SD = 1.08$). At the short interpair interval, performance on the modality questions ($\bar{X} = 1.91$; $SD = 1.41$) was superior to that on pair questions ($\bar{X} = 2.19$; $SD = 1.58$). The triple Set × Interpair Interval × Question Type interaction, F (1, 56) = 4.68, $p < .05$, summarizing these effects, is presented in Fig. 2.

At .75-sec. interpair interval, the pair set was generally disruptive, both pair questions ($\bar{X} = 2.50$; $SD = .94$) and modality questions ($\bar{X} = 2.27$; $SD = 1.55$) resulting in more errors when S was set to view the items as pairs, compared to when the modality method of storage was employed (pair question mean = 1.88; $SD = 1.98$; modality question mean = 1.55; $SD = 1.18$). The pair set was effective, however, in improving performance at the long interpair interval, pair questions having a mean of 2.22 ($SD = .88$) compared to a modality question mean of 3.33 ($SD = 1.44$); the modality set was shown to aid recall of modality questions ($\bar{X} = 2.13$; $SD = 1.16$) more than pair questions ($\bar{X} = 2.38$; $SD = 1.25$). A main effect of shift versus nonshift lures was also found, F (1, 56) = 10.99, $p < .01$, greater interference resulting from nonshift lures ($\bar{X} = 2.46$; $SD = 1.30$) than from shift lures ($\bar{X} = 2.11$; $SD = 1.47$).

Discussion

The probe-recognition procedure restored the usual effect of stimulus rate: pair questions produced more errors at the fast rate, while modality questions resulted in more errors at the slow rate. These results definitely support the hypothesis that the attenuation of the interaction between stimulus rate and recall order found in 8-item recall (Preliminary Exp.) was due to retrieval processes rather than to an alternation in the manner of item storage. When recognition only is required the interaction appears.

The modality instructional set was found to be effective in improving *Ss'* performance on modality questions, while the pair set was shown to be generally disruptive for both pair and modality questions. An examination of the triple interaction, Set × Interpair Interval × Question Type, revealed the debilitating effect of pair set to be specific to the .75-sec. interpair interval. The *Ss* instructed to view items as pairs at the fast rate were apparently unable to do so, and this attempted strategy worsened their performance on both pair and modality questions. The modality set, the "natural" method of storage at the fast rate (Broadbent, 1957), improved performance on modality questions. Pair questions were also more accurately recognized when S was instructed in the modality set than in the pair set condition. This result could be due to S's storing items as modality sets and readily retrieving analogous items from the two modalities to constitute a pair. Presumably, such a process is more easily accomplished than encoding the stimuli as pair sets at the fast rate.

At the slow rate, however, both instructional sets appeared to affect storage. Pair set improved performance on pair questions, as predicted, while modality questions were disproportionately difficult to remember under the pair set. Conversely, even though pair recall has typically been shown to be the easier recall method at the slow rate (Madsen et al., 1970, Exp. VII; Senf & Rollins, 1971), when *Ss* were instructed to store in modality sets their recall of modality questions improved while they were less able to respond accurately to pair questions.

To summarize, although an instructional set can be used to alter the encoding of items at the long interpair interval, i.e., items can be accurately stored as modality sets as well as cross-modality pairs, items cannot be efficiently stored as pairs at the short interpair interval. Therefore, while pair order retrieval of items can be prompted by induced sets at the fast rate (Senf et al., 1967), the actual manner of encoding the items does not appear to be affected.

The main effect of stimulus material found in the preliminary experiment, letters more difficult to recall than digits, was not evidenced in this experiment. The absence of any effect of materials using the probe-recognition paradigm suggests that the difficulty in recalling letters relative to digits was due to retrieval rather than storage processes.

The significant main effect of improved performance on shift items compared to nonshift items is easily explained assuming that the modality of an item is a highly salient and, therefore, frequently coded characteristic of a stimulus. The *S*s could apparently identify an item as intrusive more readily when it was drawn from a modality different from that of the item they were attempting to recognize.

Though the data are consistent with the position that the difficulty associated with recall of letters works through retrieval processes, the recognition procedure of Exp. I not only modified the recall procedure of the preliminary experiment but also changed the number of items involved in the response from eight to two. The second experiment is like the preliminary one, requiring recall, but is likewise similar to Exp. I in that only two items are required in recall. If the effects attributed to retrieval processes by virtue of their dependence on the recognition procedure of Exp. I are due to the shift from recall to recognition and not due simply to the shift from an eight- to a two-item response, the data from the *recall* probe procedure of Exp. II should mirror the results of the preliminary experiment.

EXPERIMENT II

Method

Subjects.—The 60 *S*s were students enrolled in elementary psychology classes at the University of Iowa. An equal number of males and females were randomly assigned to treatments, but sex was not analyzed as a variable. No *S* had previously participated in a bisensory memory experiment.

Apparatus and stimulus materials.—All materials were the same as those of Exp. I except that the probe-recognition cards were replaced by probe-recall cards. Probe-recall cards contained a single grid spatially depicting the eight bisensory items. On each trial, two of the item boxes were outlined in black ink to indicate which two items *S* was to recall. The probes were keyed to the probe locations tested by the recognition procedure in Exp. I.

Procedure.—The only difference between Exp. I and II procedure was the obvious change in the instructions necessitated by the shift from the probe-recognition response cards of Exp. I to the probe-recall response cards of Exp. II. Aside from the deletion of the lure-type variable, the second experiment utilized the same design as Exp. I.

Results

The means reported are average errors per *S*, maximum = 8.0. An error was scored for each stimulus item absent from the response or not assigned to the proper numerical position (grid) appropriate to its actual presentation position. Recall of letters ($\bar{X} = 2.82$; $SD = 1.17$) was significantly more difficult than recall of digits ($\bar{X} = 2.07$; $SD = 1.45$), $F(1, 56) = 27.39$, $p < .001$. A main effect of question type resulted, $F(1, 56) = 7.52$, $p < .01$, pair questions ($\bar{X} = 2.57$; $SD = 1.37$) producing more errors than modality questions ($\bar{X} = 2.32$; $SD = 1.36$). Interpair interval $F(1, 56) = 5.12$, $p < .01$, was found to affect recall, the .75-sec. interpair interval ($\bar{X} = 2.13$; $SD = 1.04$) generating fewer errors than the 2-sec. interpair interval ($\bar{X} = 2.76$; $SD = 1.58$). No other effects were statistically significant.

Discussion

The probe-recall procedure of the second experiment restored the stimulus material effects found in the preliminary experiment. While the recognition procedure, presumably eliminating or minimizing retrieval processes, found letters and digits of equal difficulty, the recall procedures requiring either eight or two digits found letters more difficult to recall than digits. Thus, these data generally support the proposition that stimulus-material effects appear to act through retrieval as opposed to storage processes.

The absence of set effects, particularly the interaction with interpair interval and type of question (pair vs. modality), most likely reflects the general attenuation of these variables' effectiveness in the eight-stimulus situation similar to that found in the preliminary experiment. More than likely, the attenuation is due to retrieval difficulty, as supported by the findings of Exp. I that the recognition procedure, minimizing retrieval demands, resulted in rate, set, and question-type effects.

The following conclusions appear warranted: First, letters are no more difficult to remember (store) than digits, in line with Miller's (1956) thinking that both types of items, though different in information value, represent similar units. However, outputting units of higher information value under high-load conditions appears to be more difficult. Second, the rate of stimulus presentation highly determines the manner of storage at the fast rate while *S*'s set can override the rate effect at slow stimulus rates. Finally, the role of both rate and set in determining the manner of storage appears to be masked when retrieval is accomplished under conditions of high information load.

REFERENCES

Broadbent, D. E. The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, 1954, **47**, 191–196.

Broadbent, D. E. A mechanical model for human attention and immediate memory. *Psychological Review*, 1957, **64**, 205–215.

Bryden, M. P. Order of report in dichotic listening. *Canadian Journal of Psychology*, 1962, **16**, 291–299.

Kintsch, W. Recognition and free recall of organized lists. *Journal of Experimental Psychology*, 1968, **78**, 481–487.

Madsen, M. C., Rollins, H. A., & Senf, G. M. Variables affecting immediate memory of bisensory stimuli: Eye-ear analogue studies of dichotic listening. *Journal of Experimental Psychology*, 1970, **83** (3, Pt. 2).

Miller, G. A. The magical number seven, plus or minus two: Some limits on the capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

Murdock, B. B., Jr. Modality effects in short-term memory: Storage or retrieval? *Journal of Experimental Psychology*, 1968, **77**, 79–86.

Senf, G. M. Development of immediate memory for bisensory stimuli in normal children and children with learning disorders. *Developmental Psychology*, 1969, **1** (6, Pt. 2).

Senf, G. M., & Feshbach, S. The development of bisensory memory in culturally deprived, dyslexic, and normal readers. *Journal of Educational Psychology*, 1970, **61**, 461–470.

Senf, G. M., & Freundl, P. C. Memory and attention factors in specific learning disabilities. *Journal of Learning Disabilities*, 1971, **4**, 94–106.

Senf, G. M., & Rollins, H. A. Effect of precuing versus postcuing retrieval order on the accuracy of bisensory memory. *Journal of Experimental Psychology*, 1971, **88**, 12–15.

Senf, G. M., Rollins, H. A., & Madsen, M. C. Recall order of dual-channel input as a function of induced recall strategy and stimulus rate. *Proceedings of the 75th Annual Convention of the American Psychological Association*, 1967, **2**, 57–58. (Summary)

Treisman, A. M. Strategies and models of selective attention. *Psychological Review*, 1969, **76**, 282–299.

(Received March 24, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 345–353

DISTRIBUTION OF REPEATED AND NONREPEATED TARGET ELEMENTS AND SHORT-TERM RECOGNITION MEMORY[1]

JOHN C. JAHNKE[2]

Miami University

Strings of digits (Exp. I) or letters (Exp. II) near span in length were presented sequentially at a 2/sec rate and were followed after brief, filled intervals by binary recognition tests of individual elements and of pairs of elements. Memory was better for single digits than pairs and for pairs of repeated digits than nonrepeated digits occupying corresponding positions in control strings. The more remote the occurrences of a pair of elements, the poorer the recognition, regardless of whether the items of the pair were different (Exp. I) or the same (Exp. II).

Norman and Wickelgren (1965) presented a target set of five different digits, an interpolated interference set of eight digits, and then gave a binary recognition test comprised of a single digit or a pair of digits. If the probe was a pair of digits, *S* was to indicate whether the digits were adjacent in the target set in the order given or in reverse order, whether only one digit was in the target set, or whether neither was in the target set. Memory operating characteristics (MOCs) of the data averaged over all serial positions showed that performance was better for single digits than pairs; furthermore, the MOC for single digits was symmetrical about the negative diagonal, whereas that for pairs was asymmetrical. The different forms of the MOCs were taken as evidence that recognition memory was different for single digits and pairs. Norman and Wickelgren pointed out (1965, p. 486) that the recognition of ordered pairs involved memory for both items and their order, while the recognition of unordered pairs could result from the independent recognition of each digit. For a variety of reasons, however, the conclusion was reached that recognition of a pair did not proceed from the independent recognition of each digit; rather, data were interpreted as indicating that *S*s coded and rehearsed the digits in sets of nonoverlapping pairs. Stated otherwise, it was presumed that some pairs were not attended to during presentation and, as a consequence, the strengths of these pairs were not increased.

Donaldson and Glathe (1969) employed a method similar to that of Norman and Wickelgren (1965), except that probe pairs were always composed of digits that had been adjacent in the target set; *S* was to indicate only whether the digits were in the original or reversed order. Again memory was better for single digits than for pairs. Pronounced serial position functions were obtained and, more importantly, the MOCs for both single digits and pairs were symmetrical. Donaldson and Glathe (1969, p. 559) suggested that their test procedure for pairs removed a confounding of order and item information inherent in the Norman and Wickelgren procedure. For the Donaldson and Glathe data, at least, there is no reason to suppose that memory for single digits is different from memory for (the order of) pairs. In particular, the retention of order information for pairs appears not to require an interpretation in terms of rehearsal of nonoverlapping pairs such as that advanced by Norman and Wickelgren.

---

[1] The data of Exp. I were collected while the author held a Special Research Fellowship, 1-F3-MH-35, 679-01, from the United States Public Health Service at the Human Performance Center, the Department of Psychology, University of Michigan. Further support was received from Grants GB-8106 and GB-28702X from the National Science Foundation to Miami University. Robert G. Crowder made many helpful comments on an earlier draft of the report of Exp. I.

[2] Requests for reprints should be sent to John C. Jahnke, Department of Psychology, Miami University, Oxford, Ohio 45056.

The two experiments to be reported here are similar to, and, in part, follow from, the studies described above. Near span-length strings of either digits or letters were presented and were followed by the shadowing of a small number of letters or digits, respectively. Binary recognition tests were given both for single items and pairs. However, the pairs were either adjacent or nonadjacent members of the target set, and *S*'s task was merely to indicate whether or not the items were, or were not, *both* members of the target set. Thus, while *S* might retain order information about the target set, this was not necessary to performance on the task, which provides a measure of item information. Furthermore, the target sets were composed either of all different elements or, alternatively, the elements were all different but one, which was repeated but once.

EXPERIMENT I

It was the intent in Exp. I to determine recognition rates for: (*a*) individual elements, and pairs of nonadjacent repeated elements and corresponding nonrepeated elements, and (*b*) pairs of nonrepeated elements as a function of the spacing of the items within the target set. It is already known that when binary recognition tests are given for single elements, a repeated element is better recognized than is the average of the nonrepeated elements, occupying corresponding serial positions (SPs) in control strings (Wolf & Jahnke, 1968). Experiment I extends that research by asking whether the pair of repeated elements is also better recognized than the corresponding control pair. Further, Exp. I provides information on the recognition of pairs of different items as a function of the number of other items intervening between the members of the pair in question (the lag of the pair). Studies of recognition and free recall of long lists of words and running recognition memory (e.g., Melton, 1967; Shepard & Teghtsoonian, 1961) have shown that recognition memory is poorer for a given item, the two occurrences of which are more remote from one another. Experiment I demonstrates such a relation for *different* items in memory sets near span in length.

Method

Subjects.—The *S*s were 10 paid volunteer female students drawn from the pool of such *S*s available to the Human Performance Center, Department of Psychology, at the University of Michigan. None had previously served in a study of short-term recognition memory, or of the effects of intraserial repetition on learning or memory, or in a study in which digits were the stimulus elements.

Apparatus and materials.—Stimuli were seven-digit strings presented on a bina-view display cell operated by a paper-tape reader. A total of 215 strings was presented to each *S* on each of 4 consecutive days. (There were, additionally, 10 practice trials before the start of the first session and 4 practice trials before the start of each subsequent session. Data from these practice trials were discarded.) The strings were derived from a basic set of 860, each of which was a random arrangement of seven different decimal digits. Half of the strings were then converted to experimental strings containing repeated digits by substituting the digit in SP 2 (or 3) for that in SP 5 (or 6). The remainder served as control strings (no repeated digits). The four sets of 215 were presented to each *S* in one of two different orders, which were counterbalanced across *S*s.

Each set of 215 contained five blocks of 43 strings, of which 17 were control strings and 26 were experimental. Half of the experimental strings in each block contained repeated digits in SPs 2 & 5 and half, 3 & 6. The strings were arranged such that neither a 2 & 5 nor a 3 & 6 string followed itself and such that there were no more than five consecutive experimental or control strings. The last digit of one string was never the same as the first digit of the next. A particular digit did not occupy the same SP in successive strings, and, with but a single exception because of *E*'s error, ascending, descending, or alternating (e.g., 579 or 642) series of length greater than two were excluded.

Procedure.—On each trial, *S* read aloud the digits of the string and a set of four interpolated letters (one of the permutations of the letters A, B, *C*, and D) that immediately followed. These elements appeared at a 2/sec rate without interruption on the bina-view screen. Immediately following the last interpolated letter, either a single digit or a pair of digits was displayed, still at the 2/sec rate. The *S* was to judge whether or not the probe, or the probe pair, was a member of the target set and to rate her confidence in her judgment on a 5-point scale. Intertrial interval was approximately 8 sec., during which time *S* gave her decision and confidence rating. The *S* was explicitly instructed that only when *both* members of a probe pair were elements of the target set was she to regard a pair as "old."

For half of the 34 control strings in two consecutive blocks of 43 trials, probes were old. The single elements (S_o) in SPs 1–7 were each probed once, as were all possible pairs (S_{oo}) composed of the elements in SPs 2–6. Only these pairs of probes were used in order to keep the number of experimental conditions manageable and to reduce the well-known effects of primacy and recency on recognition memory (Donaldson & Glathe, 1969). Probes were "new" for the remaining 17 control strings of the two consecutive blocks. Of these, there were 7 whose probes were single digits that had not appeared in the target set (S_n), 9 whose probes were pairs of digits, only 1 of which had appeared in the target set (S_{on}), and 1 whose probes were a pair of identical digits composed of a digit not included in the target set (S_{nn}).

For both the 2 & 5 and 3 & 6 conditions there were 26 strings in the cycle of 86 trials, and again the a priori probability that a probe was old was .5. Repeated elements were probed only once in the 26 instances, as were the pairs composed of a repeated element and each permissible nonrepeated element. It should be noted that the recognition of a repeated digit could be based on either occurrence in the target set, the one in SP 2 (or 3) or the other in SP 5 (or 6). Otherwise, the different elements in SPs 1–7 were each probed once, as were all possible pairs of elements from SPs 2–6. Probes were new for the remaining 13 strings of both the 2 & 5 and 3 & 6 conditions, and of these 13 there were 6 S_n probes, 6 S_{on}, and 1 S_{nn}. For strings where probes were of the class S_{on}, there were, in two consecutive blocks of 43 trials, 3 whose S_o digit was the repeated digit (S_{orn}) and 9 whose S_o digit was one of the remaining nonrepeated digits of the string. Where lures (catch stimuli) were probe pairs of the class S_{on}, the old member of the pair was equally often the first and the second member.

The *S* was instructed to respond to the probe stimulus as rapidly as she could, consistent with what she thought was a good degree of accuracy, by pressing one of the two appropriately labeled ("yes" or "no") keys. Left–right position of the keys was counterbalanced across *S*s but remained constant for a given *S*. The *S* was informed after each response of the correctness of her decision. Attention is drawn to the fact that if *S* reported old to S_{on} probes, she was told she was incorrect, but she was not told which member of the pair was new and which, old. Similarly, if *S* responded "old" to an S_{nn} probe, she was told she was incorrect but not whether the digit had appeared once or not at all. The *E* recorded *S*'s responses and response latencies, which were read to the nearest .01 sec. from a Standard Electric clock.

Results

The correct recognition and false-alarm rates for single digits are presented in Table 1 separately for the various experimental conditions. Corresponding data for probe pairs appear in Table 2. Table 1 shows that the correct recognition of individual elements was quite accurate and was a function of serial position and of repetition. The serial position data show some effect of primacy and a greater effect of recency; the dip in the serial position function for control strings at SP 6 might be expected if the seven digits were being grouped into two or more subsets (cf. Posner, 1964). Table 1 shows that the recognition of a repeated element in an experimental string was equivalent to, or better than, that for any other element in that string. Further, the recognition of a repeated element was better than that of either of the corresponding control elements. False recognition rates were relatively low and much the same for the three conditions of repetition structure. These data for the correct recognition of single digits confirm those presented earlier by Wolf and Jahnke (1968).

TABLE 1

Percentage of Hits and False Alarms for Individual Digits

Repetition structure	S_o							S_n
	SP 1	SP 2	SP 3	SP 4	SP 5	SP 6	SP 7	
Control	95	81	86	92	97	85	98	16
2&5	92	98	76	87	98	96	98	13
3&6	92	73	99	94	92	99	98	13

Note.—Because the SP of the repeated item on which recognition is based cannot be determined, the same recognition rate for repeated elements is shown at both relevant SPs.

Table 2 shows that the rate of correct acceptance of a probe pair in control strings was usually lower, and was never higher, than that for either corresponding single-element probe. For experimental strings, this rate was sometimes intermediate between the rates for the corresponding single digits. The correct acceptance of pairs also was very much a function of the particular items comprising that pair. A comparison of Tables 1 and 2 shows that a pair of repeated digits was less often correctly accepted as old than was a single occurrence of a repeated element; when, e.g., the target string was 1948247, the probe pair, 44, was less often accepted than was the

TABLE 2

Percentage of Hits and False Alarms for Pairs of Digits

Repetition structure	S_{oo}										S_{on}	S_{orn}	S_{nn}
	2&3	2&4	2&5	2&6	3&4	3&5	3&6	4&5	4&6	5&6			
Control	81	71	67	62	79	81	72	81	76	79	22	—	15
2&5	88	90	83	95	79	88	79	90	85	95	20	24	8
3&6	91	70	63	91	91	89	90	93	91	89	17	25	6

Note.—Because the SP of the repeated item on which recognition is based cannot be determined, the same recognition rate for pairs containing a repeated element is shown at both relevant pairs of SPs.

probe, 4. However, supplementary analysis not reported in detail here shows that this result stems entirely from the response bias of two *S*s to reject probes composed of a pair of repeated elements. For the remaining eight *S*s, performance was equally excellent for probe pairs composed of the repeated element and for the corresponding single-element probe.

It can also be seen, however, that a probe pair composed of repeated digits was more readily accepted as old than a pair of nonrepeated digits occupying corresponding serial positions in control strings. A sign test of these data (aggregated over the two loci of the pairs of critical digits) showed that performance was better for repeated-digit pairs than for control pairs for 9 of 10 *S*s, $p = .022$. These data included those of the 2 *S*s with response bias and, therefore, represent a conservative estimate of the magnitude of the effect examined here.

There are other features of interest in Table 2. For experimental strings acceptance of probe pairs as old appears to be related to the presence of a repeated element in the probe pair. Correct recognition was better for probe pairs that contained one occurrence of the repeated element (e.g., elements in SPs 2 & 3 in Cond. 2 & 5) than for pairs that excluded the repeated element. Similarly, a catch stimulus that included the repeated element was somewhat more likely to be falsely accepted as old than was one that did not include the repeated element. Least likely to be accepted as old was a pair of identical digits composed of an element that was not in the target set. Lures composed of identical digits were more readily accepted when the target string was a control string than when it was an experimental string. The data for probe pairs described above suggest that most *S*s almost always knew what was the repeated digit in experimental strings. The observed data then would follow, since these *S*s would be expected to reject correctly lures containing an erroneous repetition.

It may also be seen in Table 2 that the recognition of a pair of digits was a function of the lag of these digits. The data for control strings provide the clearest evidence on this matter, since the effect of repetition is not superimposed upon them. For ease of inspection the data for control strings are presented in graphic form in Fig. 1. The performance on a pair of digits is shown at the point representing the mean SP of that pair; separate functions are given for the pairs of different degrees (*d*) of lag (*d* = 0, 1, 2, or 3 for

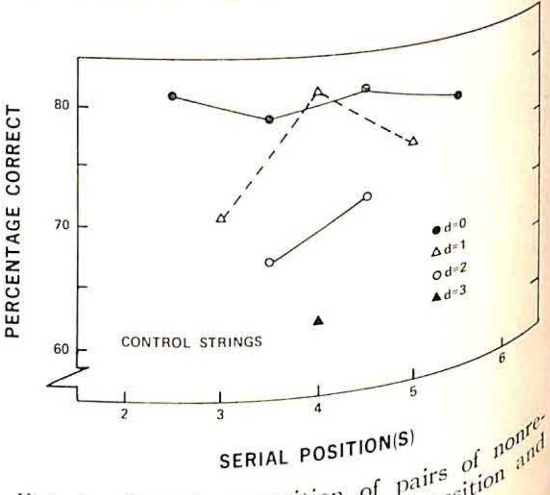


Fig. 1. Correct recognition of pairs of nonrepeated digits as a function of serial position and lag.

pairs such as 2 & 3, 2 & 4, 2 & 5, and 2 & 6, respectively). As can be seen in Fig. 1, adjacent items ($d = 0$) are recognized better than are nonadjacent items ($d > 0$), and less remote nonadjacent items are recognized better than are the more remote. The one apparently discrepant point, which could reflect sampling error, is that for the probe pair composed of items in SPs 3 & 5, for which performance is as good as for pairs composed of adjacent items. A related-measures analysis of variance was performed on the data representing mean performance for each S at each degree of remoteness. Performance was better at lower degrees of remoteness, F (3, 27) = 6.20, $p < .01$. A Newman-Keuls analysis showed that performance at $d = 0$ and $d = 1$ differed from that at $d = 3$, $p < .01$ and .05, respectively.

In Fig. 2 are presented the a posteriori probability functions for single digits and all pairs of digits combined across all SPs and conditions of repetition structure. These probabilities were based on data from S_0 and S_n probes for single digits and S_{oo}, S_{on}, and S_{orn} probes for pairs. Adjacent categories on the confidence scale were grouped, since some categories (e.g., +1, −1) were used very much less frequently than the others. The curves of Fig. 2 appear to be continuously decreasing functions; there is no indication of any threshold below which Ss are unable to order their confidence ratings. Similar functions were obtained for adjacent and nonadjacent pairs and for the three conditions of repetition structure when plotted separately.

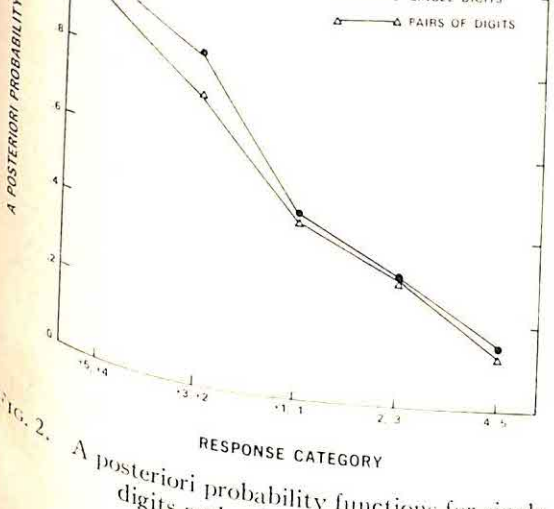


FIG. 2. A posteriori probability functions for single digits and pairs of digits.

Response latencies.—Analysis of means of median reciprocal latencies for single-digit probes showed that response speeds were faster for repeated elements than other (nonrepeated) elements in those strings and faster than for elements occupying corresponding serial positions in control strings. The serial position function for control strings provided little or no evidence of bowing; performance was poorest for the first element in the string and was best for the last.

Reaction times for probe pairs represent the time from stimulus presentation to the point at which S made her recognition response to the pair as a pair. However, for each condition for repetition structure, responses were only very slightly slower, if slower at all, to digit pairs than to corresponding single digits. Responses to probe pairs composed of the two occurrences of the repeated digits were the fastest (in 3 & 6 strings), or next to fastest (in 2 & 5 strings), of all in those experimental conditions, respectively, and were faster than to corresponding pairs in control strings.

Response speeds were examined as a function of remoteness in control strings. At $d = 0$, speeds increased regularly with mean SP of the pair. Mean performance on all other pairs was about equivalent to that for adjacent pairs from the terminal positions in the list, except for one pair (SPs 3 & 5), where responding was more rapid than for any other pair.

DISCUSSION

The present study has shown that a repeated digit was better recognized than either of the digits occupying corresponding serial positions in control strings. Further, when presented as a pair, repeated digits were more readily recognized than was the corresponding control pair. Thus, it appears that the better recognition of a repeated digit than of the average of the corresponding control digits (Wolf & Jahnke, 1968) did not arise artifactually as a consequence of the recognition of the repeated digits' being based on the trace strength of the strongest of its two occurrences or of the averaging procedure for control digits. It seems likely that repetition of an item serves to isolate that item and improve, thereby, its memorability relative to control items in the present experiment.

Of interest is the fact that correct recognition of a pair of different elements was a function of the remoteness of the elements from one another. This result is independent of the effects of serial position, since it holds when remoteness is determined with SP of either the first or last member of the pair held constant (see again Fig. 1). It apparently is the case, therefore, that memory traces are stronger in some sense for adjacent than nonadjacent pairs, and that the greater the remoteness, the weaker is the trace.

Another point of interest was the finding that the a posteriori probability functions were smooth, decreasing functions for both individual and pairs of digits. This finding suggests that there is no threshold below which responses cannot be ordered for either single digits or pairs of digits. To this degree, there is no discontinuity between the traces for either one- or two-element probes. It should be noted, however, that the influential data of Norman and Wickelgren (1965) showed a discontinuity in the a posteriori probability function (and also the MOC) for pairs of digits, but not for single digits. This discontinuity was interpreted there as showing that memory for pairs was different from memory for single digits, presumably because *S*s rehearsed the digit series as a set of nonoverlapping pairs. Present data show no need for either supposition. The reason for the discrepancy between the Norman and Wickelgren (1965) data and the present data is not clear, since there were a number of procedural differences, including, particularly, that both adjacent and nonadjacent pairs were tested in the present study. Further, the present procedure presumably tapped memory for item information alone, whereas both item- and order-information presumably were tapped by the Norman and Wickelgren procedure.

EXPERIMENT II

As noted earlier, it is well established that the recognition of a repeated element as old is poorer the greater the number of other items that intervene between the two presentations of the item (e.g., Shepard & Teghtsoonian, 1961). Experiment I demonstrated an analogous effect of lag for probe pairs composed of different items in a situation in which *S* was required to recognize whether both items of a pair were members of the memory set. It was the purpose of Exp. II to use again span-length strings and the recognition technique of Exp. I to confirm a lag effect when probe pairs were composed of two occurrences of the same item.

Method

Subjects.—The *S*s were 40 undergraduate students, 20 men and 20 women, enrolled in the introductory psychology course at Miami University. None of the *S*s had previously served in any study of short-term memory.

Apparatus and materials.—The apparatus was that of Exp. I. The target stimuli, however, were eight-consonant strings, each of which contained a single repetition of a single letter. Repeated elements occurred equally often in SP 2 and either SP 3, 4, 5, 6, or 7, or in SP 7 and either SP 6, 5, 4, or 3. There were, therefore, two instances of each of four lag conditions (lag = 0, 1, 2, or 3 different items intervening between the two occurrences of the repeated letter) and one instance of the fifth lag condition (lag = 4); one each of these nine conditions was represented in each successive block of nine trials. These choices provide two sequences of lag lengths, one of which increases systematically in the forward direction from SP 2 and the other in the backward direction from SP 7. In this way the number of potentially interfering items prior to the first occurrence of the repeated element and the number subsequent to the second, respectively, are held constant. Initially the elements of each string were drawn without known bias and without replacement from the set, CFGHJKLNQRTX, chosen so that its members were minimally phonemically confusable among themselves. The experimental strings were then derived by substituting another occurrence of the letter in either SP 2 or SP 7 for the item in the appropriate position.

Procedure.—As in Exp. I, *S*s shadowed the target string as it appeared on the bina-view screen at the rate of 2 letters/sec. Each consonant string was followed by the shadowing of a set of 4 decimal digits drawn at random from the set, 1–9 (also presented at the 2/sec rate), and by a probe pair of letters. There were four kinds of probe pairs, two of which were to be identified by *S* as old, and two, new, respectively: (*a*) *CR probes*, two occurrences of the letter that was repeated in the target string; (*b*) *CN probes*, the letter that was repeated and one of the nonrepeated letters drawn at random from those in SPs 2–6; (*c*) *ER probes*, two occurrences of one of the letters that was presented once in the target string in one of the SPs 2–6; and (*d*) *EN probes*, the letter that was repeated in the target string and a letter that was not included in the target set. For CN and EN probes, the order of presentation of the two letters was randomly determined; these letters, then, were sometimes in the order in which they were presented in the string and sometimes, not. Each of these four kinds of probe pairs was associated once with each lag condition in consecutive blocks of 36 trials. There was a total of 180 trials (strings) in the experiment proper, and these strings were presented to each *S* in one of two different orders. These trials were preceded by a practice set of nine, the data from which were discarded. The primary purpose of the CN and EN probes was to force *S* to attend to all items of the string. The use of CR and ER probes only would have permitted *S* to attend to, and rehearse, only the repeated elements.

The *S* was instructed to identify as old any probe pair, both members of which had been included in the target set, without regard to the order in which they had occurred. All other probes were to be identified as new. The *Ss* were to rate on a 4-point scale their confidence in each decision. All other details of procedure, except that the lateral position of the response keys was not perfectly counterbalanced across *Ss*, were the same as for Exp. I.

Results

Presented in Fig. 3 is the percentage of correct recognition of repeated elements as a function of lag. Separate functions are presented for the two lag conditions, one of which always included an item in SP 2 (Cond. 2) and the other, SP 7 (Cond. 7). Both functions show that the shorter the lag, the better the recognition performance. Related-measures analyses of variance performed separately for Cond. 2 and 7 showed that the effect of lag was significant, $F(4, 156) = 6.38$ and 21.16, respectively, $p < .01$, in both cases. A Newman-Keuls analysis of the data of Cond. 2 showed that performance at Lag 0 was better than at Lags 2, 3, and 4, $p < .01$, in all cases, and that performance at Lag 1 was better than at Lags 2, 3, and 4, p at least $<.05$ in each case. For Cond. 7, a Newman-Keuls analysis showed that performance at Lag 0 was better than at Lags 1, 2, 3, and 4, $p < .01$, and that performance at Lag 1 was better than at Lags 2, 3, and 4, p at least $<.05$. Figure 3 also shows that performance was consistently better in Cond. 7 than Cond. 2 at each lag; the more recent, on the average, the two occurrences of the repeated element are, the better the performance at any given lag. These data provide the basic results of Exp. II, and confirm, this time for repeated elements, the decreasing correct recognition with increasing lag observed for nonrepeated elements in Exp. I.

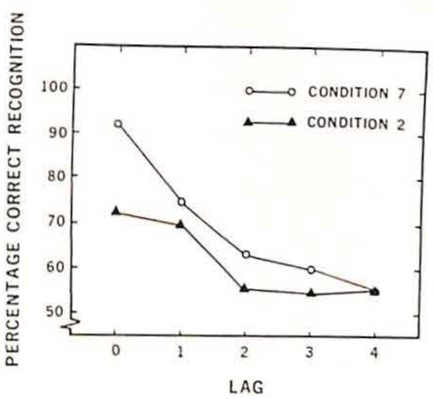


FIG. 3. Correct recognition of pairs of repeated letters as a function of lag.

TABLE 3

VALUES OF *d* COMPUTED FOR THE CONDITIONS OF EXP. II

Repetition structure	Lag				
	0	1	2	3	4
2	1.11	.94	.94	.77	.82
7	1.83	1.48	.94	1.05	.82

The question may be raised, however, whether or not the data of Fig. 3 involve differential response bias. MOCs were plotted, therefore, according to the procedures given by Norman and Wickelgren (1965). Probes of Type ER (probe pairs of identical items, item actually occurred once in the target set) supplied the data for false-alarm rates. The MOCs were well fitted by straight lines (by eye) on normal-normal probability paper. At Lag 0, the slopes were close to 1; however, the greater the lag, the less the slope. The *d* values obtained from the intersection of each MOC with the negative diagonal are shown in Table 3. These data are in general agree-

TABLE 4

Proportion of "Old" Responses for Pairs of Nonidentical Letters

Repetition structure	Probe	Lag				
		0	1	2	3	4
2	CN	.82	.78	.81	.76	.79
	EN	.26	.50	.30	.26	.27
7	CN	.78	.80	.78	.76	.79
	EN	.35	.45	.38	.37	.27

ment with those of Fig. 3: (*a*) performance is best at Lag 0 and decreases, although with some irregularity, with increasing lag; and (*b*) for any given lag, the later in the list the occurrence of the pair, the better the performance.

The proportion correct acceptance of Type CN probes (a pair of items contained in the target set, one of which was the repeated letter) and the erroneous acceptance of Type EN probes (a pair composed of the repeated letter in the target set and a letter not contained in the target set) are shown separately in Table 4 for each condition of repetition structure and lag. These data show, as might be expected, that the acceptance of CN probes was essentially constant across lag condition in both Cond. 2 and 7, in which the acceptance rates were quite similar. A comparison of the entries for CN and EN probes at each lag shows that *S*s were biased in favor of falsely accepting EN probes as old. The bias was, however, relatively small, except perhaps for pairs at Lag 1 in both Cond. 2 and 7 (items in SPs 2 & 4 and 5 & 7, respectively). The data of Table 4 also suggest that the false acceptance rates were higher for the pairs of Cond. 7. These data are presented for completeness. It will be recalled that the primary purpose of CN and EN probes was methodological—to force *S* to attend to elements of the target set in addition to the repeated elements.

Response latencies.—Reciprocals of median reaction times of correct responses to CR probes (repeated elements in the target set) were determined for each *S* for each lag condition separately. This transformation normalized the distributions of scores and permitted each *S* to contribute equally to the data, except that in Cond. 2 there were one, one, three, and two *S*s who failed to contribute data at Lags 1, 2, 3, and 4, respectively, and that in Cond. 7, one at Lag 2. The missing scores were estimated from each *S*'s relative performance in other lag conditions. These data then entered into separate related-measures analyses of variance for Cond. 2 and 7. Response speed was not systematically related to lag, and in neither analysis was the effect of lag significant, F (4, 156) = 1.98 and 2.18, respectively, p = .10, approximately, in both cases. However, response latencies were less, as shown by another separate related-measures analysis of variance, in Cond. 7 than 2, F (1, 39) = 5.03, $p < .05$. For this analysis, *S*s' data were aggregated over only Lags 0–3, since Lag 4 was common to both Conds. 2 and 7.

General Discussion

The typical running recognition-memory procedure (e.g., Shepard & Teghtsoonian, 1961) is to present twice each of a long list of items. As each item is presented, *S* is to identify the item as being new (on the occasion of its first occurrence) or old (on its second occurrence). This procedure asks by implication for the recognition of the paired occurrence of a given item; the probe, of course, is the item itself. Correct recognition is found to decline as lag increases. The present study, on the other hand, used a discrete, memory-span task adapted for recognition tests. Not only were the strings short, but, in the cases of interest here, the probe was a pair of items that either had or had not been included in the memory set, were either adjacent or not, and were either repeated or not. Both experiments reported here agreed in showing that the more remote the occurrences of the elements of the probe pair, the poorer the recognition, regardless of whether the items of the pair were different (Exp. I) or the same (Exp. II).

One explanation of the lag effect for separated occurrences of the same item is that the greater the number of intervening items, the greater the retroactive interference on the first occurrence of the item and the poorer the retention. This explanation suffices in the present context, but does not handle satisfactorily the known fact that while recognition accuracy decreases with increased lag, free recall *increases* with increased lag (Melton, 1967). A more satisfactory explanation in this regard is that *S* is presumed to rehearse items in subgroups during stimulus presentation. It follows that the first occurrence of a repeated item is encoded in a given context of items. Then, since the items of context are different at the locus of the critical item's second occurrence, recognition accuracy suffers. The more distant the two occurrences of the repeated item are, the more different are the two contexts in which the critical item is encoded and the more inaccurate is the recognition of the critical item as old (Madigan, 1969; Melton, 1967). Unfortunately, the lag effect observed in Exp. I. of the present study seems not easily accommodated by such an interpretation, since the critical items themselves are different items.

The present study also bears on an analysis of the Ranschburg effect (RE) offered previously (Jahnke, 1969; Wickelgren, 1965, 1966). It has been suggested that the RE, an inhibitory effect associated with the recall of repeated elements in a string of otherwise different elements, stems in part from *S*'s failure to recognize the repetition of an element during stimulus presentation. In this way, a bias against repetition would exist during recall. The results of Exp. II show that *S*'s ability to recognize a repeated element as repeated declines as a function of lag, even in strings near span in length, and are results quite consistent with this interpretation of the Ranschburg phenomenon.

REFERENCES

Donaldson, W. A., & Glathe, H. Recognition memory for item and order information. *Journal of Experimental Psychology*, 1969, **82**, 557–560.

Jahnke, J. C. The Ranschburg Effect. *Psychological Review*, 1969, **76**, 592–605.

Madigan, S. A. Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 828–835.

Melton, A. W. Repetition and retrieval from memory. *Science*, 1967, **158**, 532.

Norman, D. A., & Wickelgren, W. A. Short-term recognition memory for single digits and pairs of digits. *Journal of Experimental Psychology*, 1965, **70**, 479–489.

Posner, M. I. Rate of presentation and order of recall in immediate memory. *British Journal of Psychology*, 1964, **55**, 303–306.

Shepard, R. N., & Teghtsoonian, M. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, 1961, **62**, 302–309.

Wickelgren, W. A. Short-term memory for repeated and nonrepeated items. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 14–25.

Wickelgren, W. A. Associative intrusions in short-term recall. *Journal of Experimental Psychology*, 1966, **72**, 853–858.

Wolf, T. M., & Jahnke, J. C. The effects of intraserial repetition on short-term recognition and recall. *Journal of Experimental Psychology*, 1968, **77**, 572–580.

(Received March 31, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 354–362

IS VISUAL IMAGE SEQUENCING UNDER VERBAL CONTROL?[1]

ROBERT J. WEBER,[2] JOE KELLEY, AND SUSAN LITTLE

Oklahoma State University

Three experiments were conducted to examine the conjecture of R. J. Weber and J. Castleman in 1970 that sequencing between visual images of letters is sometimes under the control of verbal imagery (implicit speech) in the sense of saying a letter to oneself before visualizing it. Verbal control of visual image sequencing was investigated in alphabetic and word letter strings. Results with alphabetic letter strings were consistent with a verbal control process. However, results with word letter strings did not provide evidence for verbal control of sequencing, presumably because the name of a short word can serve to represent its several letters as a spatially parallel chunk in visual imagination. It is concluded that the visual imagery system has a limited capacity for spatially parallel representation. When this limited capacity is exceeded, as in alphabetic strings, there is implicit verbal control over visual image sequencing.

It has been suggested that nonverbal imagery (visual imagery would be an example) and verbal processes (verbal imagery in the form of implicit speech would be an example) are our two primary modes of symbolic representation (Bower, 1970; Paivio, 1971). In addition, one of the important problems of future research in cognitive psychology is seen as determining how visual imagery and verbal processes are related to one another (Bower, 1970; Paivio, 1971). In this paper interrelations of visual imagery and verbal imagery (implicit verbal processes) are examined. By "visual imagery" we mean the ability to deal directly with memorial representations of spatial information, and by "verbal imagery" we mean the use of memorial representations of linguistically related information in the form of implicit speech (Weber & Castleman, 1970). Both imageries may be accompanied by subjective impressions of "seeing" or "speaking," although as Paivio points out this is not necessarily the case.[3]

Experiment I deals with possible verbal control of visual image sequencing in an alphabetic list, plus a secondary question involving the degree of generality of scanning rates for different imagined letter forms. Experiment II serves as a control for response mode time, which was not considered in the first experiment. Experiment III is again addressed to the problem of verbal control of sequencing between visual images, but this time word letter strings as well as alphabet letter strings are examined. The comparison of word and alphabet strings is made on the conjecture that entire word images, rather than their individual letters, can be represented as a single chunk; i.e., several letters at a time can be generated by the word's name and then held in spatially parallel representation in visual imagination.

EXPERIMENT I

Two recent studies have measured the rate at which visual and verbal imagery occur. The first study (Weber & Bach, 1969) used entirely subjective procedures.

---

[1] The research reported herein was supported in part by a National Institutes of Health Biomedical Sciences Support Grant (administered by the Research Foundation, Oklahoma State University) and in part by a grant from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of the United States Office of Education and no official endorsement should be inferred.

[2] Requests for reprints should be sent to Robert J. Weber, Department of Psychology, Oklahoma State University, Stillwater, Oklahoma 74074.

[3] In the first author's classes over the past several years (N = about 500) about 95% of the undergraduates claim the ability to visually imagine distinctive letters of the alphabet. Fully 100% of the same students claim the ability to name letters in implicit speech.

In the visual imagery condition, *S*s were asked to visualize rapidly the successive letters of the alphabet, with their eyes closed, as they appeared one at a time flashed on an imaginary movie screen. The rate obtained was slightly more than 2 letters/sec. This rate was in marked contrast to verbal imagery rates with implicit and explicit speech, both of which were about 6 letters/sec, in close agreement with values from recent work using more objective procedures (Weber & Blagowsky, 1971). While these studies were highly suggestive in pointing to separate imagery systems for visual and verbal representations, the variability associated with measuring visual imagery rates underlined the need for more objective procedures. Hence, in another study, Weber and Castleman (1970) developed an objective procedure for dealing with visually imagined characters. In their Exp. II, lowercase printed letters of the alphabet were partitioned into two subsets, vertically large letters (b, d, f, g, h, j, . . . y) and vertically small letters (a, c, e, i, m, . . . z)[4]. The *S*s were instructed to visually imagine letters, in alphabetic order, and to classify each successive letter for its vertical height. This task insured that *S* process each letter for its spatial property of vertical size. The *S*s in fact reported using visual images of the letters in order to determine whether they were large or small characters. When required to abstract from the image and verbally report the visual property of each letter, the rate dropped to about 1 letter/sec. Experiment III was a control to determine whether *S*s could have rotely learned, during the experiment, the appropriate classification response for each individual letter. The results were strongly contrary to such an interpretation. The additional possibility that vertical height properties were mediated by, say, a rote verbal process acquired prior to the experiment did not accord with the subjective reports of visual images being generated nor with other evidence concerning fatigue ratings.

However, verbal mediation may have been involved in a more interesting sense. In Weber and Bach (1969) and Weber and Castleman (1970), the sequencing between visual images may have been under verbal control. In fact, several *S*s reported at the close of the experiment that they had to speak each letter implicitly before visualizing it. This would be consistent with the findings of Posner, Boies, Eichelman, and Taylor (1969), which indicate that *S*s can use letter names to generate visual characteristics of letters. It is not clear how general such a strategy is, nor whether it is optional or obligatory. For long serial lists like the alphabet, perhaps implicit verbal generation is needed to establish the sequential order of items, since the ordering capabilities of memory for unrelated visual materials appear to be limited in comparison to verbal materials (Paivio, 1971). Whatever the case, it seemed necessary to examine directly the possibility of verbal generation and control of visual image sequencing in ordered letter lists like the alphabet.

Examination of visual image and implicit verbal relations is made possible by having an objective indicator for imagery (Weber & Castleman, 1970) and by the fact that implicit and explicit (aloud) speech take place at about the same rate (Weber & Bach, 1969; Weber & Blagowsky, 1970). The experimental question concerns whether an aloud scan mode (saying each letter of a list before reporting its visual property) will result in the same rate of reporting visual properties as an unspecified scan (not aloud), in which *S* does not say successive letters explicitly but just reports their visual properties. If the rates are comparable for the two kinds of scan, then we would have evidence independent of subjective impressions that implicit speech can accompany the sequencing of successive visual images in the not-aloud scan. This, together with the subjective impression of having to say each letter before visualizing it, would be consistent with the notion of verbal generation and control of visual image sequencing. However, the implicit speech accompaniment might be purely ancillary, that is, without functional significance. So still another way is required to approach the question of verbal sequential control, and to this end we examine also the role of response indicators in denoting the spatial characteristics of visual images. Suppose that visual image sequencing is mediated by saying successive letters prior to visualizing them (the implicit speaking might be necessary to establish the sequential ordering of the letters). Then we would expect a spoken response indicator (YES or NO as to whether the letter has the property in question) to be produced and output before the next letter could be implicitly named and then visualized (saying both YES/NO and speaking the name of the next letter could not take place simultaneously). But if the response indicator were written (a line "/" for YES and a dot "." for NO), then it might be possible to begin saying implicitly the name of the next letter and producing its image while executing the written response classification for the preceding letter. This, of course, assumes that writing and speaking are at least partially independent of one another to the extent that *S*s are able to say one thing while writing something else.

---

[4] Some letters ("i"-NO, "t"-YES) are slightly ambiguous in some type fonts, but once they are designated by *E* they present no further difficulties (Weber and Castleman, 1970, Exp. III).

Method

Subjects.—The *S*s were 48 undergraduate volunteers who received extra course credit for their participation. There were 12 *S*s randomly assigned to each of four different between-*S*s conditions. Two additional *S*s were discarded for not following instructions and/or having extreme difficulty in completing the task.

Procedure.—Upon entering the experimental situation, *S*s were checked for their familiarity with the alphabet and given remedial instruction if necessary. All *S*s were shown an index card with lowercase, typewritten, letters of the alphabet in random sequence, and then instructed to differentiate between vertically large letters (YES responses) and those that were not vertically large (NO responses). This distinction was similar to that used in the Visual Property conditions of Exp. II and III in Weber and Castleman (1970), although YES/NO serve as the responses here rather than the LARGE/SMALL of the earlier study. During the experiment proper, *S* was to imagine visually the letters in alphabetic order and, as rapidly as possible, to classify each imagined letter as YES/NO. At the beginning of each trial, *S* started a remotely controlled Standard Electric clock when he was ready to begin and stopped it as soon as he had classified the last letter of the alphabet. The *S* did not receive feedback on his times, but speed and a low error rate were emphasized. If an error did occur, it was pointed out to *S* at the end of the trial. This constituted the Lowercase (LC) condition. In addition, there was a comparable Uppercase (UC) condition in which *S*s distinguished between letters with long vertical lines (B, D, E, F, H, I, . . .) which required YES responses and those without long vertical lines (A, C, G, J, O, Q, . . .) which required NO responses. Each *S* was to imagine both LC and UC alphabetic lists alternated over 10 blocks of LC and UC trials. The LC and UC conditions were balanced for serial position across *S*s.

There were also four between-*S* conditions. The first two involved scanning mode. An *S* scanned the alphabet by speaking it aloud (SA) or by a nonspecified, not-aloud (NA) procedure as he generated in imagination successive letters to examine and respond to their visual properties. In the SA conditions, *S* would proceed as follows (for LC letters): "a"-NO, "b"-YES, "c"-NO, . . ., "z"-NO; that is, speak each letter before responding with YES or NO to describe the vertical height of the imagined letter. In the NA conditions, no instruction was given regarding scanning, and *S* did *not* speak aloud the successive letters before responding overtly with YES or NO. That is, he would proceed as follows: NO, YES, NO, . . ., NO. It was, of course, possible that *S* would generate the successive letters in speech imagery before generating their visual images, even though he received no instruction to do so, but this was an empirical question to be answered by the similarity of processing rates for the SA and NA conditions. Two other between-*S* conditions involving spoken or written responses (SR, WR) were combined factorially with the scanning mode conditions. In the SR conditions, *S* simply said YES or NO for the visual property of each successive letter. In the WR conditions, *S* wrote a long vertical line "/" corresponding to a YES response or wrote a dot "." corresponding to a NO response. The *S* was to classify the alphabetic string as rapidly as possible and to keep his eyes closed, in all conditions, so he would not visually monitor his WRs.

In summary, the design consisted of a factorial arrangement of 2 Scanning Modes (SA, NA) × 2 Response Classes (SR, WR) × 2 Letter Cases (LC, UC), with scanning mode and response class between *S*s and letter case within *S*s. The principal response measure was time to process the 26 member imagined alphabetic list.

Results and Discussion

Descriptive findings are presented in Fig. 1, in which mean time in seconds is shown as a function of practice. The left ordinate shows mean time per 26 letters,

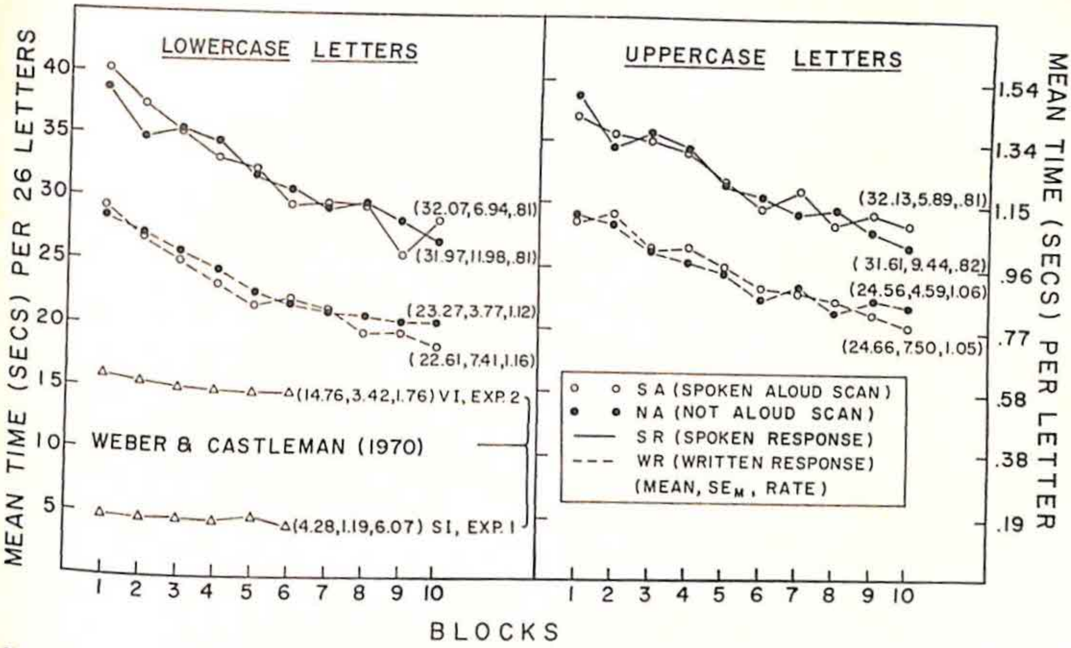


FIG. 1. Response time as a function of scanning mode, response mode, and blocks. (Left panel is for lowercase letters and right panel for uppercase letters.)

and the right ordinate is the same information on a time-per-letter basis. The left-hand panel is for lowercase letters and the right-hand panel, for uppercase letters (the bottom half of the left panel contains results from previous experiments that will be commented on later). The parenthetical number triple next to each function represents the mean, standard error of the mean, and the scan rate in letters per second processed. Each mean is based on 120 observations (12 *Ss* × 10 Blocks); the standard error of the mean is a between-*Ss* measure of variability based on each *S*'s mean; and the scan rate is simply the reciprocal of the mean. Errors (omission and commission) occurred for fewer than 2% of the letters (a phase error was counted only once). Those trials on which errors did occur were not temporally disparate from correct trials, so no further distinction is made.

Mean processing times are almost identical for spoken scanning (SA) and unspecified scanning (NA), with F (1, 44) < 1. The absence of a significant scan effect is consistent with the view that implicit speech scanning is involved in the NA conditions. However, the differences in variability for the two scans do suggest at least some difficulty with assuming that completely identical processes are involved.

There is a substantial difference between WR and SR modes, F (1, 44) = 14.55, $p < .01$. Employing a WR mode takes much less time than an SR mode, even though writing is probably a slower response system than speaking. Again this is consistent with the view that sequencing between visual images is accompanied by implicit verbal processes. SRs appear to delay the hypothesized verbal control processes involved in generation and sequencing between successive visual images of letters.

Finally, the results for upper- and lowercase letters suggest that visual imagery rates are about the same for two quite different classes of visual forms, F (1, 44) = 2.56, $p > .05$. This would tend to increase the generality of prior findings (Weber & Bach, 1969; Weber & Castleman, 1970) by suggesting that visual imagery rates are constant over fairly broad fluctuations in visual form. Just how large a variation in visual form would be possible before differences in rate occurred is open to question. None of the main effects showed significant interactions with one another.

Large practice effects are also evident in Fig. 1. This is in contrast to some other imagery tasks, as shown in the bottom half of the left panel. The visual imagery (VI) function is from Weber and Castleman (1970, Exp. 2) and did not require *S* to abstract the height properties of the letters or to respond YES/NO. The procedure was entirely subjective; *S* simply stopped the clock when he had completed visualizing, one at a time, the letters of the alphabetic string. The large practice effects with the visual property task of the present study would seem, then, to be a result not of visual image representation but of abstraction and/or overt response time. Both response conditions of the present study seem to be stabilizing somewhat above the VI condition, as would be expected if abstracting height properties and responding YES/NO takes additional time over visualizing letters. The bottom function is for verbal or speech imagery (SI) and also from Weber and Castleman (1970, Exp. 1). It simply required *S* to say to himself, as rapidly as possible, the letters of the alphabet and stop the clock when finished. It provides an estimate of the minimum component time that would be required of a verbal control letter-naming process in a VI condition. Hence, if the last five blocks of the visual property written conditions of this study are compared with the VI and SI functions, it is suggestive that the combined VI and SI times are close to the time for the written conditions.

In reviewing the evidence for verbal control in the sequencing of images, perhaps most compelling is the multiple approach to the verbal control issue. The *absence* of a significant scanning effect (SA, NA) is consistent with an implicit speech letter-by-letter naming process in the NA conditions. The *presence* of a significant response mode effect (SR, WR) is consistent with an implicit speech process, generating and controlling the sequencing of letters, which is delayed by the overt verbal response of categorizing letters as YES/NO. And once again, there is the strong subjective impression of having to say each letter before visualizing it.

A cautionary note is still in order. The response mode effect would come as no surprise if writing were in fact a more rapid response mode than speaking. This unlikely possibility is considered in Exp. II. Another related possible explanation of the response mode effect is that the written line and dot responses are initially more compatible with the height properties of visual images than are the spoken YES/NO responses. But if this were the case, we would expect WR and SR mode functions to converge rapidly across blocks of practice, something that does not occur.

Each of the findings and arguments for verbal control may not be completely compelling in its own right, but taken together they provide a strong line of converging evidence pointing to the simple notion that *S* says each letter before visualizing it. The generality of such a notion is immediately suspect, and Exp. III in this series is directly addressed to the question of when verbal accompaniment of visual image sequencing does and does not take place.

Experiment II

It seems obvious that in most instances "equivalent" materials can be spoken at least as rapidly as they can be written. The findings and interpretations of Exp. I depend on such an assumption. However, we could not seem to find evidence on this point in the recent psychological literature. Also it is not completely clear how equivalent spoken YES/NOs are to written lines and dots—perhaps the written responses are more rapid. A direct comparison of writing and speaking rates is called for, in order to strengthen the rationale of Exp. I. Such a comparison also has intrinsic interest since writing and speaking are, perhaps, our two most important means of signaling information to others about our internal states. Moreover, writing and speaking are widely used as response modes in psychological research, without much regard to their potentially very different output rates and the implications this has for, say, short-term-memory decay work.

In this experiment, then, response times for speaking and writing are assessed. For future use, two different symbolic materials are employed: an alphabetic list and a binary list consisting of spoken or written equivalents of YES/NO. The written analog of YES/NO is, as in Exp. I, a vertical line and dot, respectively.

Method

Subjects.—The *S*s were 10 volunteers from an undergraduate psychology class.

Design and procedure.—A factorial design was used, with 2 Types of String (Alphabet, Binary) × 2 Response Modes (Written, Spoken). Both factors were within *S*s. In the written condition of the alphabet task, *S*s were asked to write the alphabet in their normal connected cursive handwriting as rapidly as possible. (Response time was always the dependent variable.) In the spoken alphabet condition, *S*s were asked to speak the alphabet aloud as rapidly as possible. In the spoken binary condition, *S*s spoke aloud a 26-member series of alternating YES/NOs (YES, NO, YES, NO, . . .). In the written binary condition, *S*s wrote out a 26-member series of alternating lines and dots ("/./." etc.). The lines and dots correspond to the YES/NO written classification of Exp. I. In the binary tasks, we did not wish to require *S*s to count their responses to have only 26, so a sheet of paper was divided into seven columns 1 in. wide. For the written binary task, *S* wrote his responses in clusters of 4/column (/./.; /./.; etc.), except for the last column which required only 2 responses to equal the 26. For the spoken binary task, *S* kept track by moving his finger to a new column on completion of each cluster of four spoken responses.

Procedures for timing were the same as for Exp. I. Each *S* was given six trials in each of the four conditions. A trial consisted of a single processing, i.e., writing or speaking an alphabetic or binary list. Following the six trials of the first condition, each *S* was given the six trials of the next condition, etc. The orders of the four conditions were randomized for each *S*.

Results and Discussion

Mean response time in seconds for each task is shown in Table 1. The response time for the WR mode is greater than for the SR mode on both the binary and alphabet strings. An analysis of variance shows that the type of response has a significant effect, $F(1, 9) = 172.22$, $p < .01$. The type of string does not have a significant effect, $F(1, 9) = 1.84$, $p > .05$. However, there is an obvious significant interaction between type of response and type of string, $F(1, 9) = 99.92$, $p < .01$. The main comparison of a priori concern is that of the binary spoken and written conditions. Spoken responding is reliably faster than written responding, with correlated $t(9) = 3.87$, $p < .01$.

TABLE 1

MEAN AND STANDARD ERROR OF THE MEAN FOR PROCESSING 26 ITEM BINARY OR ALPHABETIC STRING, EXPERIMENT II

Statistic	Binary		Alphabet	
	Written	Spoken	Written	Spoken
$\bar{X}$	9.37	7.50	12.87	4.71
SE_M	1.55	1.80	2.07	1.30

Note.—Means averaged over 10 *S*s × 6 trials for each condition. SE_M is measured in seconds.

The greater response time for the WR mode as compared to the SR mode for both the binary and alphabet strings reflects in part response times for different motor acts. The significant String × Response interaction can be interpreted as meaning that with written responses, *S*s can execute the simple vertical lines and dots required by the binary task faster than the more complex curved lines required by the alphabet task. But when spoken responses were used, *S*s could say the alphabet faster than the string of 26 alternating YES/NOS.

These results suggest that the differences between WR and SR modes of Exp. I may have been conservative, since in sheer response time (unaccompanied by a visual image task and with a completely regular alternation of YES/NO responses), written responding is slower than spoken responding. It is, of course, possible to argue that sequences other than an alternating YES/NO would produce different results, that different sequential patterns would interact differently with the two response modes, but such effects, if they occur, are likely to be small indeed compared to the basic distinction between writing and speaking.

EXPERIMENT III

Experiment I indicated that sequentially generating the visual image representations of an ordered list like the alphabet is under verbal regulation. But it is clearly the case that sequential ordering for at least some visual images is not verbally controlled in the sense of first having to say a verbal counterpart of each imagined item in the list. Likely examples would include dreams, hallucinatory experiences, and visual reverie. Such images seem to have a sequential life of their own, and their content and order often defy verbal description. Another case in which individual images do not seem to be under sequential verbal control would seem to be spatially parallel composite images, that is, images with several different individual components appearing simultaneously but separated in subjective space. An example would be a short word, the letters of which might be imagined at the same time. Thus the letters of the word "cat" might be represented in visual imagination sequentially "c," "a," "t," or simultaneously "cat." Extracting visual spatial properties from a letter string that is sequentially generated (alphabet) may well be different from abstracting visual spatial properties from a letter string in which all letters might be represented simultaneously in imagination, as in a short word.

In this experiment, the relative rate of abstracting visual properties from different types of letter strings, i.e., from the alphabet and from words, is measured under different scanning and response conditions. By using two different scanning modes, spoken and nonspoken, and two different response modes, written and spoken, the presence or absence of individual verbal letter generation in visual imagery for word versus alphabet strings can be investigated. If visual image formation of the successive individual letters of a word is also governed by saying implicitly each successive letter, then relative scanning times should be the same for spoken and nonspoken scanning, and written responding should be faster than spoken responding. The rationale is the same as Exp. I. However, suppose the word name (together with the appropriate instructional set) is sufficient to arouse a spatial simultaneously available visual image of the entire word (or at least several letters of it). Then speaking each individual letter of the word would not be necessary prior to visualizing its letters, and an NA scan should take less time than a letter-by-letter SA scan. Also, the difference between the SR and WR modes should diminish for the NA scan in comparison to the SA scan, since in the former condition full advantage could be taken of the simultaneously available image without the additional potentially disruptive process of saying each successive letter one at a time.

Method

Subjects.—The *S*s were 40 undergraduate college students who participated in order to fulfill an introductory psychology course requirement. Two potential *S*s were not used because they could not recite the alphabet correctly. One other potential *S* was not used because he could not learn to represent imagined large and small lowercase, typed letters of the alphabet with YES and NO, respectively. The *S*s were assigned at random to conditions, with 10 *S*s per between-*S*s condition.

Design and procedure.—The experimental tasks used in this experiment involved the use of the alphabet and 18 four-letter words; all words were of one syllable and were high-frequency-usage words (defined as A or AA by Thorndike and Lorge). Each of the words contained both vertically large and small letters, e.g., "girl." From images of the two different letter strings (words and alphabet), *S*s were to abstract visual properties of the successive letters. Thus the image "girl" would give rise to the response sequence YES, NO, NO, YES. The design and procedure were similar to that of Exp. I and consisted of a factorial arrangement of 2 Response Modes (Written, Spoken) × 2 Scans (SA, NA) × 2 String Types (Alphabet, Word). The first two factors were between *S*s and the last factor was within *S*s.

Each *S* was given several preliminary practice trials for abstracting visual properties from the letters of the alphabet and for abstracting visual properties from the letters of a practice word. A spoken ready signal was given, and *E* started the remotely controlled clock with the onset of the spoken cue ("alphabet" or "X", where X was one of the stimulus words). The *S* was to respond as rapidly as possible by classifying the successive letters as YES/NO. The *E* also stopped the clock on completion of the visual property "spelling" of the word or alphabet, i.e., after the response to the fourth or twenty-sixth letter. This was a departure from the previous two experiments in which *S*s controlled the clock. The departure was necessary because pilot data showed that *E* was more accurate than *S*s in measuring the short response times involved with words in this experiment. The response times were measured from the onset of the alphabet or word cue to the response to the last letter in a string. To provide greater accuracy in measuring response times, *E* also tape-recorded all responses and later played back the tapes at half-speed. The *S*'s spoken responses were recorded as usual. To tape-record *S*'s written responses, a soft lead pencil was used to write on an aluminum sheet to which a microphone was taped. The distinctive sounds made by *S*'s pencil on the aluminum sheet were readily picked up and served to demarcate response onsets. Again, *S* did not visually monitor his responses.

Each *S* was given 10 trial blocks. Each block was composed of 6 stimulus words of 4 letters each and one alphabet string. This allowed for roughly the same amount of work per block for both types of letter string (actually 24 letters for word strings and 26 letters for the alphabet string in a given block). The 6 words for each *S* were randomly selected from the pool of 18, 4-letter words. The same 6 words appeared in each of the 10 trials given to a particular *S*, but different *S*s received different sets of 6 words to increase generality. The order of appearance for words and the alphabet was randomized from block to block.

Results

Each *S*'s mean response time across trials was calculated. In order to compare directly *S*'s responses in processing the word and the alphabet letter strings, his response times in each condition were computed on a time-per-letter basis. Therefore, each *S*'s mean processing time per letter was the unit used in the subsequent analyses.

As can be seen from Table 2, the processing times are slightly faster in the NA scan conditions than in the SA scan conditions, $F\ (1, 36) = 5.53$, $p < .05$, with most of the variance coming from the SA word conditions. Processing times are also faster in the word condition than in the alphabet condition, $F\ (1, 36) = 97.99$; and faster in the WR condition than in the SR condition, $F\ (1, 36) = 46.01$, both $p\text{s} < .01$.

Of particular interest are the interaction effects in this experiment. There is a significant Scan × String interaction $F\ (1, 36) = 11.20$, $p < .01$, indicating the differences between spoken and nonspoken scans were greater in the word condition than in the alphabet condition. There is also a significant Response × String interaction, $F\ (1, 36) = 15.27$, $p < .01$, showing that the differences between SRs and WRs were greater in the alphabet condition than in the word condition. Finally, there is a significant three-factor Response × Scan × String interaction, $F\ (1, 36) = 11.20$, $p < .01$, indicating the relatively unique character of word strings with SRs and NA scanning.

TABLE 2

MEAN, STANDARD ERROR OF THE MEAN (SE_M), AND RATE AS A FUNCTION OF STRING, SCAN, AND RESPONSE MODE, EXPERIMENT III

String and scan	Response mode					
	Spoken			Written		
	M	SE_M	Rate	M	SE_M	Rate
Alphabet						
SA	1.01	.14	.99	.61	.16	1.64
NA	.96	.27	1.04	.60	.15	1.67
Words						
SA	.86	.10	1.16	.49	.12	2.04
NA	.53	.09	1.89	.47	.12	2.13

Note. Means are determined by averaging over 10 *S*s and 10 trials per condition. SE_M is measured in sec/letter and rate is measured in letters/sec.

Practice effects across trial blocks warrant comment. They were relatively flat for word strings, with no indication of a Condition × Practice interaction. The alphabet strings had practice effects similar to the corresponding conditions of Fig. 1, Exp. I, again with no indication of a Condition × Practice interaction.

Discussion

The equivalence of SA and NA scanning for alphabetic strings agrees with the findings of Exp. I and of Weber and Blagowsky (1970). The present results are also consistent with the notion that the visual imaging of individual letters in an alphabet string is under verbal control in the sense that each letter is implicitly spoken prior to imagining it. In contrast, the results for word strings are not equivalent for spoken and nonspoken scanning and are therefore inconsistent with the notion that *S* implicitly speaks each letter of a word when using a nonspoken scan mode. Verbal control over the sequential generation of visual images of the letters is not needed with the word letter-strings used in this study. Instead, on the basis of a word's spoken name, *S* may visually generate the word as a whole, or at least with more than one letter at a time simultaneously represented in visual imagination. The *S* could then scan left to right for the visual properties of the simultaneously available letters without having to fall back on implicit verbal processes. This is indeed the subjective impression of what is done. The time advantage with nonspoken scanning would arise because the names of the individual letters constituting the word would not need to be verbalized, since the visual image representation of the word would be simultaneously available prior to scanning individual letters. Hence, with word strings the verbal YES/NO responses should be comparable in rate to the written YES/NO equivalents, as is the case. Along these same lines, the sequential generation problem disappears, even for alphabet strings, if the string is perceptually available. Weber and Kelley (1972) have found that the spoken/written difference for the visual property task disappears when the alphabet string is visually present in front of *S*. They hypothesize that the sequential ordering problem is handled in that case by left-to-right eye movements over the string rather than through the use of implicit verbal processes as suggested here.

While the processes suggested for word and alphabet strings in the present series of experiment may seem dichotomous, it is likely that scanning a long polysyllabic word for the visual properties of its letters would be part way between the limiting cases of short familiar words versus long serial alphabetic lists. We would expect spatially parallel visual image representation to be possible within familiar letter clusters with a one- or perhaps two-syllable name, but sequencing between syllables for longer words might well be under verbal control, as with the alphabetic strings. Thus the visual property scanning of the imagined lowercase word "oklahoma" might allow for the spatially parallel visualization of letters within a syllable, while the sequencing between syllables would involve verbal processes. This would certainly seem to be the case with a word like "supercalifragilisticexpialidocious."

The picture that emerges is of a visual imagery system with a capacity for limited, spatially parallel representation (Paivio, 1971). When a familiar letter string with a short name falls within that capacity, visual image sequencing is not verbally generated. But when a letter string exceeds that capacity, sequential generation and ordering become necessary and visual imagery comes under the control of verbal imagery in the form of implicit speech.

REFERENCES

BOWER, G. Analysis of a mnemonic device. *American Scientist*, 1970, **58**, 496–510.

PAIVIO, A. *Imagery and verbal processes.* New York: Holt, Rinehart & Winston, 1971.

POSNER, M. I., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph*, 1969, **79** (1, Pt. 2).

WEBER, R. J., & BACH, M. Visual and speech imagery. *British Journal of Psychology*, 1969, **60**, 199–202.

WEBER, R. J., & BLAGOWSKY, J. Metered memory search with implicit and explicit scanning. *Journal of Experimental Psychology*, 1970, **84**, 343–348.

WEBER, R. J., & BLAGOWSKY, J. Metered memory search and concurrent chanting. *Journal of Experimental Psychology*, 1971, **89**, 162–170.

WEBER, R. J., & CASTLEMAN, J. The time it takes to imagine. *Perception & Psychophysics*, 1970, **8**, 165–168.

WEBER, R. J., & KELLEY, J. Aspects of visual and acoustic imagery. *Psychonomic Science*, 1972, **27**, 121–122.

(Received March 31, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 363–370

SERIAL AND PARALLEL ENCODING PROCESSES IN MEMORY AND VISUAL SEARCH[1]

JANE MARANTZ CONNOR[2]

Vanderbilt University

In two experiments using an S. Sternberg paradigm, test-set size and delay between presentation of memory and test sets were varied. The existence of a serial memory-set encoding stage was inferred from the observed effect of memory-set size on reaction time (RT) over and above its contribution to the number of comparisons to be performed. In Exp. I, in which digits were used, items in the test set appeared to be encoded in parallel, as the only contribution of test-set size to RT was in the number of comparisons to be made. In Exp. II, the auditory and visual confusability of the items in both sets was varied. Evidence of a serial test-set encoding stage was obtained at the shorter delays.

Several lines of research indicate that visually presented material is transformed as it is stored in memory. Sperling (1960) and Averbach and Coriell (1961) used a partial report technique to show that tachistoscopically presented items are available to *S* in a visual sensory information store that decays in about .5 sec. when the display is followed by a bright postexposure field. Using a "same–different" procedure, Posner and Keele (1967) estimated that the visual information from a single letter decays in about 1.5 sec, when this information is not relevant to task performance.

Other experiments suggest that short-term memory (STM) is auditory in character. Using written recall of visually presented letters, it has been shown that auditorily similar letters are more likely to be confused than visually similar letters (Conrad, 1964; Wickelgren, 1965). More errors are made also in the recall of auditorily similar lists (Conrad, Baddeley, & Hull, 1966; Conrad & Hull, 1964). Incorporating these findings, Sperling (1963, 1967) proposed a model for STM based on the conversion of the rapidly fading visual image into an auditory image that is subsequently maintained by rehearsal.

Experiments in memory and visual search have generally ignored the properties of the encoding process, which transforms a visually presented memory item into an STM representation. In search studies, *S*s typically have several seconds to study the memory list before the test items are presented. Sternberg (1967b) found that reaction time (RT) in this paradigm is a linear function of the number of comparisons to be performed between memory and test items. If s is the number of items in the memory set and d the number of items in the test set, then negative responses are predicted by the function

$$\mathrm{RT} = a + bsd, \qquad [1]$$

in which b is the mean time per comparison and a is the time required for all other processes.

Varying the visual clarity of a single test stimulus, Sternberg (1967a) established the existence of a test-stimulus encoding stage. The clarity of the test stimulus increased RT, but had virtually no effect on the rate of memory search. This result and others led Sternberg (1969) to propose the four-stage model of memory search shown in Fig. 1.

EXPERIMENT I

The first experiment sought evidence for the existence of an additional memory-set encoding stage by varying the amount of

---

[1] The author wishes to thank Robert C. Calfee and Edward E. Smith for their guidance in the completion of this research. Part of the results in the report were presented at the meeting of the Western Psychological Association, Los Angeles, April 1970.

[2] Requests for reprints should be sent to Jane M. Connor, Vanderbilt University, Department of Psychology, Nashville, Tennessee 37240.

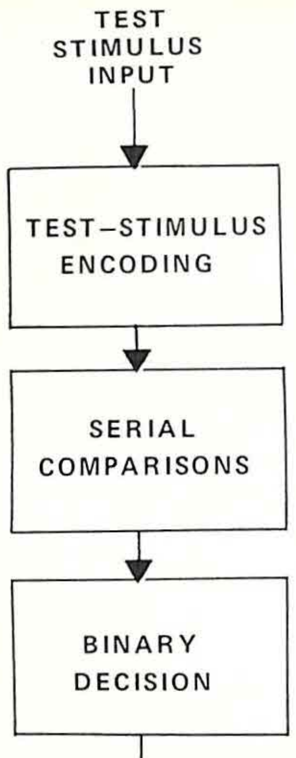


FIG. 1. Model of memory search proposed by Sternberg (1969).

delay between presentation of the memory and test sets. It was assumed that the encoding of the memory set requires time proportional to the number of items contained therein. When a short delay separates the presentation of the memory and test sets, the encoding of the memory set cannot be completed prior to the presentation of the test items, and hence additional processing time is required in proportion to the number of memory items. For negative responses, RT is predicted to conform to the following relation:

$$\mathrm{RT} = a + cs + bsd, \qquad [2]$$

in which a, b, s, and d are as previously defined and c, the mean additional time necessary to complete the encoding of each item in the memory set, should be inversely related to the amount of the delay.[3] (The data for positive responses were analyzed for linearity and consistency with the negative responses. However, the unsettled question of a self-terminating or exhaustive search makes it uncertain how many comparisons are performed for positive responses. Thus parameter estimates and model fitting were restricted to the negative responses.)

Method

Design and procedure.—Three delays between offset of the memory set and onset of the test set (150 msec., 500 msec., and 2,000 msec.) were combined factorially with four levels of memory-set size (1, 2, 3, and 4), four levels of test-set size (1, 2, 3, and 4) and two response types (positive and negative). The *Ss* served for 10 sessions and received all 96 conditions during each session. The data from the first session were considered as practice and were not included in the analysis.

The memory set was presented for 200 msec., followed by a postexposure field for the varying delay (thus yielding interstimulus delays of 350 msec., 700 msec., and 2,200 msec.) and then the test set for 2 sec. The luminance of all three fields was adjusted to 30 ftl.

Each session was divided into three blocks of 32 trials, delay being fixed in each block and order of delays counterbalanced across days and *Ss*. The *Ss* were instructed to respond as soon after the onset of the test set as possible but to maintain high accuracy.

Stimuli.—Six sets of 32 pairs of cards were prepared, each set containing all combinations of test-set size, memory-set size, and response type. The stimuli in the test and memory sets were randomly drawn from the digits 0–9 (Transartype A-59, dry transfer). On positive trials, a single target appeared in the test set. The serial position of the target item in both sets was randomly determined. There were no repeated digits in either set.

The items in the memory set were arranged in a horizontal row 1.3° above the fixation point; items in the test set were in a horizontal row 1.3° below the fixation point. Each digit was 15′ high and a maximum of 12′ wide. A complete set of four digits subtended 1.5°. Four fixed positions were defined in the horizontal row for each set. When a set contained only one digit, one of the two middle

---

[3] Briggs and Blaha (1969) report a similar relationship in an experiment in which the memory set was kept constant in each session. The parameter c was interpreted as reflecting memory retrieval time.

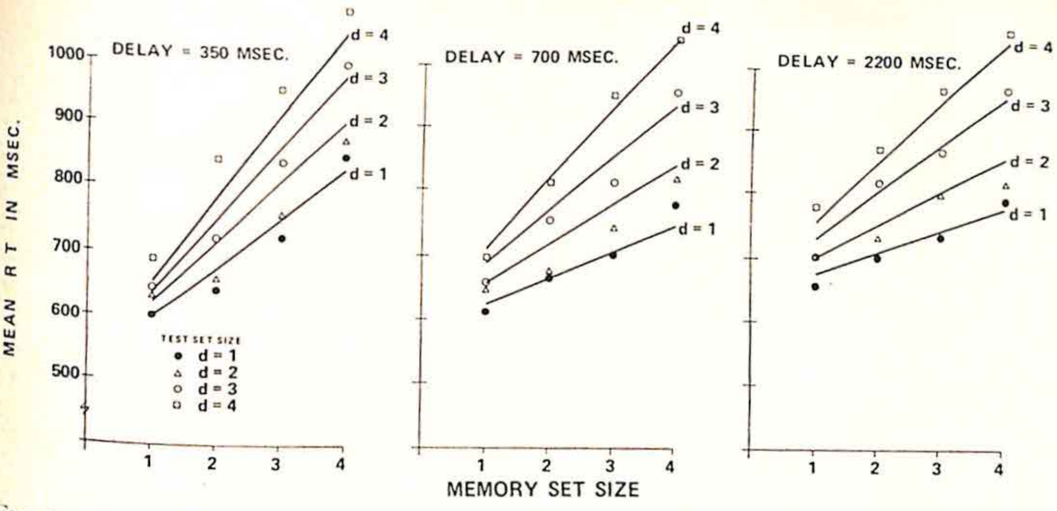


FIG. 2. Exp. I. Mean RT for negative responses as a function of memory-set size, test-set size, and interstimulus delay. (Lines are based on Equation 2 and Table 2.)

positions was used, for two digits the two middle positions were used, for three digits two middle and one end position(s) were used, and for four digits all positions were used.

Apparatus.—The stimuli were presented to *S* in a Scientific Prototype (Model GB) three-channel tachistoscope. A switch activated by *S* initiated a trial and triggered a Standard Electric timer accurate to the nearest .01 sec., which stopped when a voice relay was activated by the verbal response (either "Yes" or "No") of *S*.

TABLE 1

SIGNIFICANT SOURCES OF VARIANCE AND ORTHOGONAL TREND COMPONENTS, EXPERIMENT I

Source	*df*	Sums of squares	*F*
Response (R)	1	16997	72.3
Delay (D)	2	3148	10.2
Test set size (T)	3	43089	81.4
Linear	1	41591 (96.5%)	
Nonlinear	2	1498 (3.5%)	
Memory set size (M)	3	75275	101.1
Linear	1	75230 (99.9%)	
Nonlinear	2	45 (.1%)	
T × R	3	1775	12.1
Linear × R	1	1429 (80.5%)	
Nonlinear × R	2	346 (19.5%)	
M × R	3	1992	32.8
Linear × R	1	1714 (86.0%)	
Nonlinear × R	2	278 (14.0%)	
M × T	9	5745	36.3
Linear × Linear	1	4741 (82.5%)	
Linear × Nonlinear	2	123 (2.1%)	
Nonlinear × Linear	2	500 (8.7%)	
Nonlinear × Nonlinear	4	381 (6.7%)	
M × D	6	2849	10.1
Linear × D	2	2485 (87.2%)	
Nonlinear × D	4	364 (12.8%)	
M × T × R	9	1454	6.4
Linear × Linear × R	1	608 (41.8%)	
Linear × Nonlinear × R	2	336 (23.1%)	
Nonlinear × Linear × R	2	173 (11.9%)	
Nonlinear × Nonlinear × R	4	337 (23.2%)	

Note.—For source of variance, $p < .01$.

Subjects.—The *S*s were seven male and five female undergraduates paid for participating.

Results

Mean RT of the correct responses was computed for each *S* and experimental condition and submitted to an analysis of variance. (The overall error rate was 3.1%, with only slight variations across experimental conditions.) The significant sources of variance and the orthogonal trend components are shown in Table 1. Of particular interest was the finding of a significant interaction of delay with memory-set size and with no other experimental factor. Rate of memory search increased with increasing delay, whereas rate of visual search was unaffected. This result was taken as reflecting the existence of a serial memory-set encoding stage, and

TABLE 2

ESTIMATES OF MEAN ADDITIONAL TIME PER MEMORY ITEM FOR COMPLETION OF ENCODING (c), MEAN TIME PER COMPARISON WITH A SINGLE TEST ITEM (b), AND INTERCEPT (a), EXPERIMENT I

Delay	c	b	a
350 msec.	58.0	18.5	520
700 msec.	19.2	24.5	570
2200 msec.	7.2	23.1	647

Note.—Estimates are recorded in milliseconds.

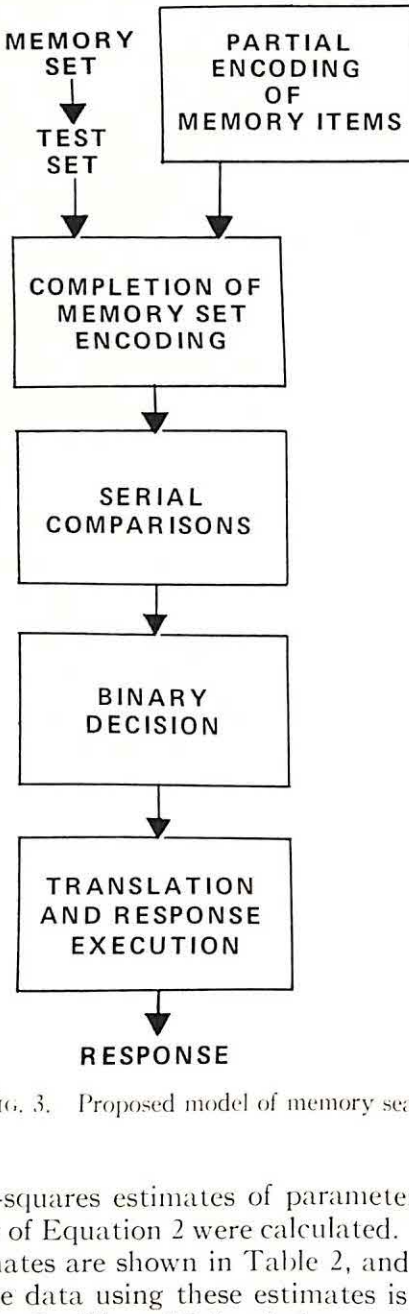

MEMORY SET
TEST SET
PARTIAL ENCODING OF MEMORY ITEMS
COMPLETION OF MEMORY SET ENCODING
SERIAL COMPARISONS
BINARY DECISION
TRANSLATION AND RESPONSE EXECUTION
RESPONSE

FIG. 3. Proposed model of memory search.

least-squares estimates of parameters *a*, *b*, and *c* of Equation 2 were calculated. These estimates are shown in Table 2, and the fit to the data using these estimates is shown in Fig. 2. The additional time required to encode the memory items, *c*, decreased with increasing delay, whereas the time taken for each memory- and test-set comparison, *b*, was 6 msec. less at the shortest delay, but otherwise invariant with delay. The inincrease in parameter *a*, the time taken for all processes other than comparisons and encoding, with increasing delay is interpreted as the result of the increasing temporal uncertainty in the three delay conditions.

Estimates of mean comparison time, *b*, were obtained for each level of memory-set and test-set sizes after subtracting the appropriate estimate of mean encoding time (Table 3). An analysis of variance of these comparison time estimates indicated that the average estimate obtained from the memory search functions (22.7 msec.) was less than that of the visual search functions (26.2 msec.). Subsequent tests on the significant Type of Search (Memory or Visual) × Numbers of Items in the Set interaction also indicated that mean comparison time estimated from memory search functions was greatest when there was one item in the test set and shortest when there were two.

Discussion

The results of this experiment are interpreted in terms of the model shown in Fig. 3. This model differs from Sternberg's (1967a) (see Fig. 1) only in the first stage. This difference reflects differences in the nature of the experimental manipulations. The existence of a serial memory-set encoding stage in the present experiment was inferred from the observed decrease in rate of memory search, with decreasing delay between the presentation of the memory and test sets. It is postulated

TABLE 3

ESTIMATES OF MEAN COMPARISON TIME OBTAINED FROM MEMORY SEARCH AND VISUAL SEARCH FUNCTIONS, EXPERIMENT I

Delay	Memory search		Visual search	
	d	Mean comparison time	*s*	Mean comparison time
350 msec.	1	23.9	1	30.1
	2	15.0	2	32.1
	3	20.7	3	26.4
	4	18.2	4	21.7
700 msec.	1	30.2	1	19.9
	2	19.4	2	24.3
	3	25.7	3	26.5
	4	24.8	4	22.8
2200 msec.	1	28.4	1	33.0
	2	17.2	2	26.5
	3	25.8	3	25.7
	4	22.7	4	25.5

Note.—Estimates are recorded in milliseconds.

that memory comparisons are performed on the output of the encoding stage, i.e., on the representations of the targets in STM. Otherwise, both memory and visual search would be affected equally as delay was varied; in fact, only memory-set size and delay interacted. The encoding of the test items (Sternberg, 1967a) in this experiment was apparently performed in parallel since the main contribution of the number of test items to RT was in the time required to make the necessary comparisons between memory and test items (Equation 2).

However, it is not clear that all of the encoding of the memory items must be carried out prior to the start of the comparisons. The possibility of some time sharing of the two processes is indicated by the faster comparison time at the shortest delay in which the encoding time was greatest (Table 3). This hypothesis is further supported by the finding that mean comparison time was greater when estimated from memory search than from visual search functions. Each additional memory item required both extra encoding and comparison time but the extra comparison time was not quite so much as that required by an additional test item.

EXPERIMENT II

The purpose of the second experiment was to investigate further factors affecting the encoding and comparison stages in memory and visual search. Several *E*s have been concerned with effects of auditory and visual confusability on memory and visual search processes. As opposed to the strong effect of auditory confusability obtained in recall tasks (Conrad, 1964; Conrad et al., 1966), visual confusability appears to be a more important factor in search tasks than auditory confusability (Chase & Calfee, 1969; Chase & Posner, 1965; Krueger, 1970; Neisser, 1963). Neisser (1963) proposed that comparisons of visual features formed the basis of visual search or scanning performance. The smaller effects of auditory confusability in search tasks, then, might not be directly due to the comparison process but rather to characteristics of the encoding process. If, in the process of forming memory representations, *S*s automatically enrich the stimulus by adding auditory information, then auditory confusability would be expected to increase encoding time but not necessarily comparison time. Thus it was predicted that visual confusability would increase comparison time while auditory confusability would increase encoding time.

Method.—Three levels of interstimulus delay, 350 msec., 700 msec., and 2,200 msec., (during which the memory set was available for 200 msec.) were factorially combined with three levels of memory-set size (1, 2, and 4), three levels of test-set size (1, 2, and 4), three types of material (auditorily confusing, visually confusing, and neutral), and two response types (positive and negative). The auditorily confusing letters (B, C, P, E, T, V, Z, D) were chosen from Conrad's (1964) error matrix. The visually confusing set (B, C, D, G, O, Q, R, U) was defined by an a priori overlap measure (Chase & Posner, 1965). These sets of letters and the neutral set (A, D, H, I, M, O, Y, Z) were closely equated for relative frequency of occurrence in English and naming time of the letters. Letters in the memory and test sets were drawn at random from the appropriate set for each trial.

Each of 12 *S*s served for 10 sessions, each session containing one observation under all 162 conditions. The first session served as practice. Trials were blocked by delay, and within a block all three types of material were randomly presented. The apparatus, experimental procedure, and design of the cards were otherwise identical to Exp. I.

Results.—Mean RT of correct responses was computed for each *S*, and condition and submitted to an analysis of variance. (Overall error rate was 5.2%. The small variations in relative frequency of errors across conditions paralleled the RT variations.) Linear components accounted for a large percentage of the variance in the data both in the main effects of memory-set size and test-set size and the higher order interactions of these factors (Table 4). Thus a linear model was fitted to the data. Unlike Exp. I, however, it was necessary to include a linear component of test-set size in addition to the linear components of memory-set size and the product of the two. (Note the significant Memory-Set Size × Test-Set Size × Response Type × Delay interaction in this experiment.) For negative responses, the following model was fitted to the data using least squares procedures

$$RT = a + cs + ed + bsd. \qquad [3]$$

TABLE 4

SIGNIFICANT SOURCES OF VARIANCE AND ORTHOGONAL TREND COMPONENTS, EXPERIMENT II

Source	*df*		Sums of squares	*F*
Material (Ma)	2		6244	24.8
Response (R)	1		80170	94.7
Test set (T)	2		168,178	103.7
Linear		1	167,664 (99.7%)	
Nonlinear		1	514 (.3%)	
Memory set (Me)	2		238,024	151.3
Linear			237,608 (99.8%)	
Nonlinear			416 (.2%)	
Ma × Me	4		1203	3.9
Ma × Linear			1097 (91.2%)	
Ma × Nonlinear			106 (8.8%)	
R × T	2		17,667	28.9
R × Linear			17,549 (99.3%)	
R × Nonlinear			118 (.7%)	
R × Me	2		8230	21.2
R ×Linear			7277 (88.4%)	
R × Nonlinear			950 (11.6%)	
Delay (D) × T × Me	8		3269	5.5
D × Linear × Linear		2	2846 (87.0%)	
D × Linear × Nonlinear		2	301 (9.2%)	
D × Nonlinear × Linear		2	104 (3.2%)	
D × Nonlinear × Nonlinear		2	18 (.6%)	
Ma × T × Me	8		1938	4.5
Ma × Linear × Linear		2	563 (29.1%)	
Ma × Linear × Nonlinear		2	413 (21.3%)	
Ma × Nonlinear × Linear		2	464 (23.9%)	
Ma × Nonlinear × Nonlinear		2	498 (25.7%)	
D × R × T × Me	8		2251	5.4
D × R × Linear × Linear		2	1692 (75.2%)	
D × R × Linear × Nonlinear		2	107 (4.8%)	
D × R × Nonlinear × Linear		2	315 (14.0%)	
D × R × Nonlinear × Nonlinear		2	137 (6.0%)	

Note.—For sources of variance, $p < .01$.

in which *s* equals memory-set size and *d* equals test-set size. The parameter *b* was interpreted as mean comparison time, *c* as mean additional encoding time per memory item, *e* as mean encoding time per test item, and *a* as the time for all processes other than encoding and comparisons. These parameters were estimated separately for each material type and delay using least-squares procedures (Table 5). Mean RT as a function of memory-set size, test-set size, delay, and material type is shown in Fig. 4. The lines in Fig. 4 were drawn using Equation 3 and the estimates in Table 5.

Effects of material type appeared only at the longer delays where, relative to the neutral material, an increase in encoding time was observed for the auditorily confusing material and an increase in comparison time for the visually confusing material. As opposed to the relative independence of the encoding and comparison parameters in Exp. I, a clear trade-off relationship was obtained in Exp. II. Mean encoding time decreased, while mean comparison time increased with increasing delay. At the longest delay, mean comparison time for the neutral letters was similar to that obtained by Sternberg (1967b).

DISCUSSION

The results of these two experiments confirm the existence of a serial encoding process operating on the memory set in a recognition memory experiment. This is reflected in a linear increase in RT as a function of memory-set size over and above the time taken to compare memory- and test-set items when only a brief interval separates the presentation of the memory and test sets. The prediction that auditory confusability would increase encoding time and visual confusability would increase comparison time was confirmed in Exp. II. However, the evidence for a serial test-set encoding process and the trade-off relationship between encoding and comparison parameters in Exp. II indicates that the relatively straightforward model depicted in Fig. 3 is not completely adequate. It is not considered reasonable to assume that comparisons between memory- and test-set items required only 6–7 msec. at the shortest delay. Rather, it is inferred that some of these comparisons were completed during the 80 msec. per memory- and

TABLE 5

ESTIMATES OF MEAN COMPARISON TIME (*b*), ADDITIONAL ENCODING TIME PER MEMORY ITEM (*c*), ENCODING TIME PER TEST ITEM (*e*), AND INTERCEPT (*a*) AS A FUNCTION OF TYPE OF MATERIAL AND DELAY, EXPERIMENT I AND EXPERIMENT II

Material type	Delay	*b*	*c*	*e*	*a*
Auditorily confusing letters	350	8.3	81.6	70.4	472
	700	14.8	63.2	58.3	477
	2200	33.6	34.5	28.0	598
Visually confusing letters	350	5.9	84.5	87.7	462
	700	28.7	52.8	51.3	508
	2200	53.6	3.5	−2.0	690
Neutral letters	350	6.3	86.5	77.2	435
	700	14.5	50.6	56.3	511
	2200	38.6	13.6	−1.1	650
Digits (Exp. I)	350	18.5	58.0	.2	520
	700	24.5	19.2	−.6	570
	2200	23.1	7.2	.1	647

Note.—Estimates are recorded in milliseconds.

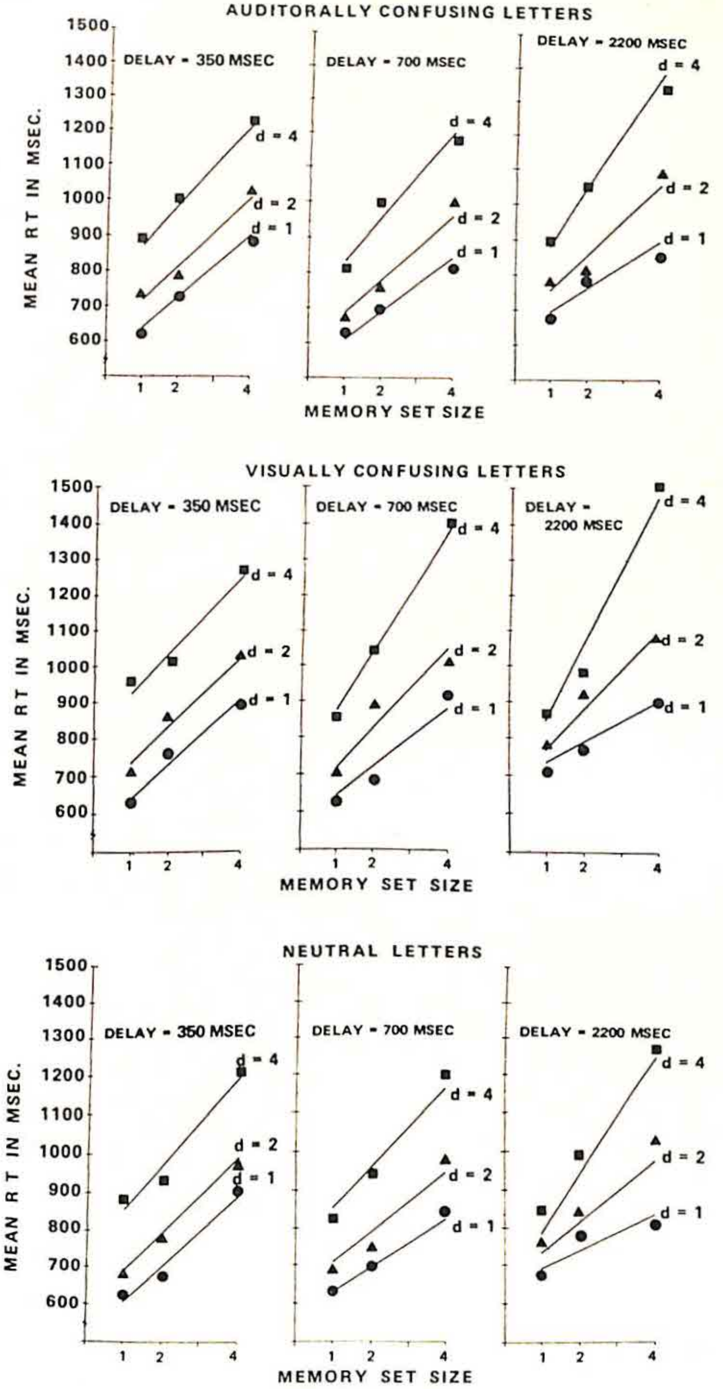


FIG. 4. Mean RT for negative responses as a function of memory-set size, test-set size, interstimulus delay and material type, Exp. II. (Lines are based on Equation 3 and Table 5.)

test-item additional encoding time. With increasing delay additional encoding time decreased and thus the additional time necessary to complete the requisite comparisons increased. Thus given the relative invariance of comparison time across delay in Exp. I, it may be justifiable to interpret the estimates of c as the requisite additional encoding time per memory item. The estimates of comparison and encoding time obtained in Exp.

II, however, may not be so literally construed. The data strongly imply the existence of a serial memory-set encoding process as well as a serial comparison process, but the durations of these processes may not be strictly independent.

The finding of a serial test-set encoding process in Exp. II was unexpected. Sternberg (1967a) established the existence of a test-set encoding process. But as he used only one item in the test set, there is no evidence for this being either a serial or a parallel process. At the present time, it appears that test items are encoded serially when both *S*'s processing capacity is absorbed with the encoding of other relevant information (i.e., the memory set) and when the two sets of material are highly confusable. The first condition derives from the decrease in test item encoding time with increasing delay. At the longest delay, the visually confusing and neutral test sets were encoded in parallel. The second conclusion comes from comparing the two experiments and other findings that indicate the discriminability of items in a display may determine whether it is processed serially or in parallel (Connor, 1972). That the results with the neutral displays followed the same pattern as the results with the confusable ones in Exp. II rather than the digits in Exp. I is not surprising. It is possible that the neutral letters were more confusable than the digits. More likely it is probable that with each *S* receiving each type of display in a mixed-list design *S*s were unable to employ different types of processing strategies for the different types of displays. There is evidence that a discrimination performed in parallel in one situation may be performed serially when embedded among more difficult discriminations requiring serial processing (Connor, 1972).

REFERENCES

Averbach, E., & Coriell, A. S. Short-term memory in vision. *Bell System Technological Journal*, 1961, **60**, 309–328.

Briggs, G. E., & Blaha, J. Memory retrieval and central comparison times in information processing. *Journal of Experimental Psychology*, 1969, **79**, 395–402.

Chase, W. G., & Calfee, R. C. Modality and similarity effects in short-term recognition memory. *Journal of Experimental Psychology*, 1969, **81**, 510–514.

Chase, W. G., & Posner, M. I. The effect of auditory and visual confusability on visual and memory search. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, April 1965.

Connor, J. Effects of increased information load on the parallel processing of visual displays. *Perception & Psychophysics*, 1972, **12**, 121–128.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

Conrad, R., Baddeley, A. P., & Hull, A. J. Rate of presentation and the acoustic similarity effect on short-term memory. *Psychonomic Science*, 1966, **5**, 233–234.

Conrad, R., & Hull, A. J. Information, acoustic confusion and memory span. *British Journal of Psychology*, 1964, **55**, 429–432.

Krueger, L. E. The effect of acoustic confusability on visual search. *American Journal of Psychology*, 1970, **83**, 389–400.

Neisser, V. Decision-time without reaction time: Experiments in visual scanning. *American Journal of Psychology*, 1963, **76**, 376–385.

Posner, M. I., & Keele, S. W. Decay of visual information from a single letter. *Science*, 1967, **158**, 137–139.

Sperling, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, **74** (11, Whole No. 498).

Sperling, G. A model for visual memory tasks. *Human Factors*, 1963, **5**, 19–31.

Sperling, G. Successive approximations to a model for short-term memory. *Acta Psychologica*, 1967, **27**, 285–292.

Sternberg, S. Two operations in character recognition: Some evidence from reaction time measurements. *Perception & Psychophysics*, 1967, **2**, 45–53. (a)

Sternberg, S. Scanning a persisting visual image versus a memorized list. Paper presented at the meeting of the Eastern Psychological Association, Boston, April 1967. (b)

Sternberg, S. The discovery of processing stages: Extensions of Donders' Method. In W. G. Koster (Ed.), Attention and Performance II. *Acta Psychologica*, 1969, **30**, 276–315.

Wickelgren, W. A. Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, 1965, **70**, 102–108.

(Received April 1, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 371-379

CARDIAC AND RESPIRATORY ACTIVITY DURING VISUAL SEARCH[1]

MICHAEL G. H. COLES[2]

University of Illinois

Measures of cardiac and respiratory activity were recorded from human *Ss* while they performed a series of visual search tasks which varied in target stimulus discriminability. In Exp. I, tonic heart rate was found to be lower and cardiac deceleration greater for low- as compared to high-discriminability tasks. Exp. II replicated the major findings of Exp. I and, in addition, showed that the duration of respiratory period decreased less from an alert to a task level, for low- as opposed to high-discriminability tasks. The results for heart rate were interpreted as supporting J. I. Lacey's intake-rejection hypothesis rather than an effort hypothesis.

Lacey and his co-workers (Lacey, 1959, 1967; Lacey, Kagan, Lacey, & Moss, 1963) reported that the direction of change in heart rate (HR) from an alert to a task situation was dependent on the nature of the task to which *S* was exposed. Cardiac deceleration was observed for tasks such as empathic listening and visual attention, while other tasks, such as reverse spelling and mental arithmetic, were associated with cardiac acceleration. Lacey and his co-workers argued that HR increase appeared for tasks which required *S* to reject the external environment, while HR decrease occurred when intake of, or attention to, the external environment was required. It was also noted that some tasks (e.g., listening to the rules of a game), designed to require both intake and rejection, were associated with no change in HR. As a result, these investigators suggested that the tasks described above could be placed on an "intake-rejection" continuum, tasks at the intake pole being associated with HR decrease, and those at the rejection pole with HR increase.

Several studies have questioned the interpretation of these phenomena in terms of the intake-rejection continuum. Campos and Johnson (1966, 1967; Johnson & Campos, 1967) concluded that the verbalization dimension was sufficient to account for the observations of Lacey and his co-workers (Lacey et al., 1963), although Edwards and Alsip (1969) have since argued that "the question of whether verbalization and intake-rejection are one and the same remains [p. 11]." It has also been suggested that Lacey's results can be interpreted in terms of a dimension of demand for action (Elliott, 1969) or of effort (Campos & Johnson, 1967). Both these and the verbalization interpretations are clearly in line with traditional arousal theory (e.g., Duffy, 1962).

A serious problem concerning the lack of an independent operational measure of the task situations has contributed to the controversy over the appropriate dimension. Only by having such a measure can the situations be objectively located on the dimension in question. Tursky, Schwartz, and Crider (1970) did use an independent measure by giving their *Ss* mental arithmetic tasks which varied in difficulty, and obtained results that supported the intake-rejection hypothesis. However, Tursky et al. were concerned with *phasic* aspects of cardiac activity and their study, therefore, contrasts with those described above where the concern was with *tonic* HR.

The present studies were designed to examine the effects of attentional demands on tonic HR in situations where such demands are objectively manipulated. This manipulation was accomplished through the use of a series of visual search tasks which varied with respect to the discriminability of target stimuli from background stimuli. It was proposed that low-discriminability tasks (searching for "e" target letters among "a" background stimuli) require more external attention or more environmental intake than high-discriminability tasks (searching for "b" target letters among "a" background stimuli), the independent measure of discriminability being provided by searching speed. In terms of the intake-rejection continuum, low-discriminability tasks can be placed nearer the intake pole than high-discriminability tasks. Accordingly, it was hypothesized that tonic HR is lower, and cardiac deceleration greater, under the low-discriminability conditions. Under an effort hypothesis, the opposite prediction would be made, since low-discriminability tasks intuitively require more effort. The experiment, therefore, also provides a test of the relative merits of the intake-rejection and effort hypotheses.

---

[1] This research was supported by a grant from the University of Illinois Research Board. The assistance of Charles Lewis, for statistical advice, and Peter Rozday and Brian Sosdian, for help in running *Ss*, is gratefully acknowledged.

[2] Requests for reprints should be sent to Michael G. H. Coles, Department of Psychology, University of Illinois, Champaign, Illinois 61820.

In order to test the effects of target-letter identification, the density of target letters was varied. A further manipulation involved the response requirements following target-letter identification in that *S*s were required to either count or check target letters. Since under counting conditions *S* was required to store and process digits, it was proposed that HR would be higher under counting, as compared to checking conditions.

Porges and Raskin (1969) found that HR variability decreased during attention-involving tasks. Thus, a final prediction was the HR variability, like tonic HR, would be lower during searching than during preceding alert and rest periods.

EXPERIMENT I

Method

Subjects.—The *S*s were 32 male undergraduate students recruited from an introductory psychology course at the University of Illinois.

Apparatus and procedure.—The experiment took place in a partially soundproofed, air-conditioned, electrically shielded room. Beckman Biopotential skin electrodes, a Beckman R-411 dynagraph, and a 4-channel Hewlett-Packard Model 3960 FM tape recorder were used to provide a continuous recording of each *S*'s heartbeats on chart paper and on analog tape.

Each *S* was given a consecutive series of 16 trials, each trial consisting of a rest period (50 sec.), an alert period (35 sec.), and a task period (35 sec.), during which *S* performed a visual search task. The task material consisted of 16, 40 × 60 letter matrices, which had been typed on stencils and then duplicated on mimeograph paper. The predominant letter on each matrix was the letter "a": One of two possible target letters was distributed in the matrix, as will be described shortly. All letters were lowercase elite type.

Each task was defined in terms of four parameters: (*a*) target letter—either "b"(B) or "e"(E); (*b*) density—either high, with 32 target letters in the matrix, or low, with 16 target letters; (*c*) response mode—either count or check; and (*d*) pattern—either Pattern 1, 2, 3, or 4.

The target letters were randomly distributed in the matrix, the different patterns being obtained by starting at different points in a random number table. Patterns 1 and 3 were different configurations of the high-density matrix, and Patterns 2 and 4, different configurations of the low-density matrix. All patterns occurred with every level of every other parameter, except density, giving 16 possible combinations of the four parameters.

The 16 tasks were presented to the *S*s in a random order with the following constraints. First, the eight possible combinations of target letter, density, and response mode had to occur within the first eight trials. Second, the same level of a particular parameter could not occur on more than two consecutive trials. Of the possible orders that resulted from these constraints, eight were used, four *S*s being given each order.

Each *S* reported individually for the experiment and was instructed to scan each matrix, line by line, beginning in the top left-hand corner, and to work from left to right along each line. He was asked to scan as fast as he could but to try to avoid missing any target letters. When he detected a target letter, *S* was instructed either to make a check mark through it with a pencil or to count it. Two practice tasks were given, after which the series of 16 trials began.

Throughout the experiment, *E* sat with *S*. This *E* received auditory signals via headphones, which served as cues for him to change from one period of the experiment to the next. Marking pulses, which occurred at the same time as the auditory cues signifying the beginning of alert and task periods, were recorded on the analog tape. Following the first cue, *E* instructed *S* to relax and placed a stack of three sheets of paper in front of him. After the next cue, 35 sec. later, *E* told *S* to get ready and removed a blank sheet of paper from the top of the stack. This revealed instructions consisting of the target letter to be expected and the response required on the forthcoming task. After the third cue, 35 sec. later, *E* told *S* to begin the task, at the same time removing the instruction sheet, which revealed the task sheet. When *E* received the cue for the beginning of the next trial, he instructed *S* to stop searching and relax. For counting conditions, *S* was also told

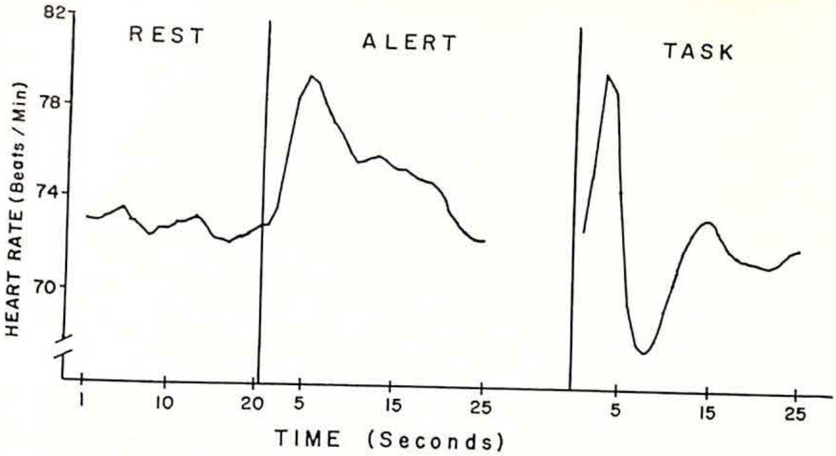


FIG. 1. Average second-by-second HR values for the three experimental periods, Exp. I.

to write down the number of target letters he had counted in the margin beside the line he had just scanned.

Scoring.—1. Speed of searching. For each *S*, on each task, a score was obtained representing the number of lines scanned. For checking tasks, this was based on the last line in the matrix on which there was a check mark. For counting tasks, the last line scanned was taken as the line next to the *S*'s written total of the number of target letters detected.

2. Heart rate. Second-by-second analysis of HR was provided by an IBM 1800 computer. For the first 25 sec. of both alert and task periods and the last 20 sec. of rest periods, R–R intervals were obtained to the nearest .004 sec. (For the rest periods, the values were obtained by playing the tape backwards.) These R–R intervals were converted to HR values, and a mathematical procedure was uesd to obtain one HR value for each second. This procedure is similar to that used in the hand-scoring method of second-by-second HR analysis, the only difference being that in the method presented here, the onset of the first complete heart beat following period onset is taken as the starting point of the analysis, instead of the actual onset of the experimental period. The HR during the three periods was only sampled since any effects due to the experimental manipulations should be evident within the sample time.

Results

Analysis of tonic HR in the three experimental periods.—Figure 1 shows the results of second-by-second analysis of HR. The phasic accelerative-decelerative-accelerative response pattern at the beginning of the task period probably reflects the effects of postural changes and instructions, which occurred when *S* began the task. The initial response pattern terminated about 16 sec. after task onset, after which HR became relatively stable. Thus, in order to determine whether there were differences between E and B tasks in tonic HR, mean HR values were computed for each *S* for each task by averaging second-by-second HR values for Sec. 16 through 25 of the task periods. The 16 values obtained were reduced to 8 by averaging across pattern, and an analysis of variance with three within-*S* factors (target letter, density, and response mode) was performed on the resulting data. As shown in Fig. 2, HR was lower during E than during B tasks, $F\ (1,\ 31) = 5.69,\ p < .05$. HR during counting tasks (71.60) was lower than during checking tasks (72.18), $F\ (1,\ 31) = 5.66,\ p < .05$. (These and all other

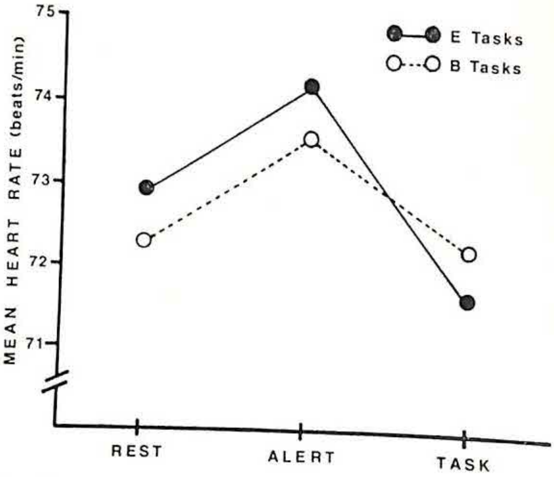


FIG. 2. Mean HR values, representing tonic HR, for the three experimental periods, Exp. I.

values of HR are in bpm.) A significant Target Letter × Density interaction indicated that while HR was lower for E than for B tasks under low-density conditions, the reverse was the case for high density.

Similar analyses were performed on the data for tonic HR during the rest and alert periods. The relevant means were based on the second-by-second HR values for Sec. 16 through 25 of alert periods and the last 10 sec. of rest periods. Mean HR during rest periods prior to E tasks was higher than that prior to B tasks, F (1, 31) = 5.01, $p < .05$. During alert periods, a similar difference was obtained, although it failed to reach significance, F (1, 31) = 3.42, $p = .07$. Both these effects can be seen in Fig. 2.

Analysis of cardiac deceleration and change in tonic HR.—For each S and each task, values representing cardiac deceleration were computed by taking the difference in mean HR between alert and task periods. The 16 values obtained for each S were reduced to 8 by averaging across pattern, and an analysis of variance of the same type as described in the previous section was performed on the resulting data. Mean HR decreased more for E tasks (−2.57) than for B tasks (−1.38), F (1, 31) = 9.79, $p < .01$, and mean HR decreased more under counting (−2.39) than under checking (−1.56) conditions, F (1, 31) = 3.96, $p = .055$. A significant Target Letter × Density interaction indicated that although cardiac deceleration was always greater for E than for B tasks, this difference was maximal under low-density conditions.

For each S, three HR values were computed representing average tonic HR for each of the three periods. There was a significant change in tonic HR across the three periods, F (2, 64) = 5.29, $p < .01$. Scheffé tests (Scheffé, 1959) indicated that HR was lower during task periods (71.89) than during alert periods (73.86). However, HR during rest periods (72.61) was not significantly different from that during either alert or task periods.

Analysis of HR variability.—Similar analyses to those performed on mean HR values for 10-sec. samples during alert, rest, and task periods were performed on the values for the standard deviation (SD) of HR for the same samples. None of the main effects or their interactions were significant in these analyses.

In order to determine whether HR variability changed across the three periods, three SD values were computed for each S representing the average SD of HR during the three periods. SD changed significantly across periods, F (2, 64) = 16.73, $p < .001$. Scheffé tests indicated that HR variability was significantly lower during the tasks (3.72) than during either alert (5.16) or rest (4.91) periods.

Analysis of searching speed.—More lines were scanned during B tasks (29.23) than E tasks (15.30), F (1, 31) = 98.84, $p < .001$, and fewer lines were scanned under checking (20.00) response conditions than under counting (24.53) conditions, F (1, 31) = 53.39, $p < .001$. A significant Target Letter × Response Mode interaction indicated that the difference in lines scanned between E and B tasks was greater under counting than checking conditions.

Heart rate and searching speed.—Each S performed each of the eight task types (combinations of target letter, density, and response mode) under two pattern conditions. By comparing the number of lines scanned for a particular task type under one pattern condition with the number scanned under the other pattern condition, it was possible to categorize each S's searching speed on each task type as either "fast" or "slow." Thus, each S's tasks were divided into a fast and slow group, each group containing a representative of each of the eight task types. The mean HR values for each S on each of the 16 tasks were used to compute average tonic HR values for each S on fast and slow tasks. Over all Ss, average tonic HR values were 71.74 and 72.06 for fast and slow tasks, respectively. This difference was not significant, $t = .86$, $p > .05$. Cardiac deceleration (the difference in tonic HR between alert and task) was greater for fast (−2.10) than for slow

(−1.85) tasks, but this difference was also not significant, $t = .43$, $p > .05$.

Over all *S*s, the mean number of lines scanned was 25.19 for fast tasks and 19.34 for slow tasks.

Discussion

In accordance with the principle hypothesis, tonic HR was lower, and cardiac deceleration greater, for E than for B tasks. Since E tasks were performed more slowly than B tasks, there is some justification for concluding that the manipulation of discriminability was successful and that E tasks required more attention. However, the differences in HR between E and B tasks cannot necessarily be attributed to attentional differences between the tasks. It could be argued that the slower HRs observed during E tasks, were due to the slower searching speed on these tasks. But, if this were the case, then a relationship between HR and searching speed, independent of task type, would be predicted. The analysis of the difference in HR between fast and slow tasks indicated that mean HR was lower and cardiac deceleration greater for fast tasks. Although not significant, these results clearly show that searching speed did not elevate HR and that the differences in HR between E and B tasks cannot be attributed to differences in searching speed.

One implication of the difference in searching speed between E and B tasks is that more identifications were made under B conditions. However, more identifications were also made under high- than under low-density conditions, since there was no effect of density on searching speed. Thus, if differences in frequency of identifications were responsible for the differences in HR between E and B tasks, then similar HR differences should be found for high- versus low-density conditions. None of the analyses indicated that this was the case. Therefore, the results indicate that the differences in HR and cardiac deceleration between E and B tasks were due to attentional differences rather than to differences in searching speed or the frequency of identifications.

Aside from the main effect of target letter, the analysis of cardiac deceleration showed a significant Target Letter × Density interaction. Since there was no difference in searching speed between high- and low-density conditions, *S*s were spending more time searching and less identifying under low-density conditions; that is, under low-density conditions searching activity is maximal. Thus, if HR is related to the attentional requirements of searching, the effects of target letter should be maximal under low-density conditions. This was precisely the case.

An unexpected finding was that E and B tasks differed with respect to the HR during preceding alert and rest periods. This may have been due to the constraints placed on the order of presentation of the tasks. These constraints resulted in the majority of E tasks being preceded by B tasks, and the majority of B tasks being preceded by E tasks. If HR tends to be high in B tasks and this effect persists through subsequent rest and alert periods, then HR during rest and alert periods prior to E tasks would be expected to be high. For similar reasons, HR prior to B tasks would be expected to be low.

The response requirements of the task had an effect on both searching speed and HR. The slower speed observed under checking conditions was probably due to the motor response required under these conditions. The same factor may have been responsible for the high HRs observed under these conditions. Any elevating effect on HR of digit processing under counting conditions was not evident from comparisons of HR under the two response requirements.

As predicted, HR variability was lower during the attention-involving tasks than during preceding alert and rest periods. Tonic HR also changed across experimental periods, being lower during the task periods than during the alert periods. However, HR was not significantly lower during the tasks than during rest periods. This latter result is not consistent with Lacey et al. (1963), who found a significant change in HR from a resting level to a lower level during an environmental intake task.

EXPERIMENT II

The data described in Exp. I suggest that tonic HR and cardiac deceleration vary as a function of the discriminability of the target letters in a visual search task. In particular, the results support the hypothesis that HR is lower and cardiac deceleration greater the more external attention that is required from *S*.

Experiment II was designed to replicate the major conditions of Exp. I and to provide additional information concerning the relationship between attentional variables and respiration. Several studies (e.g.,

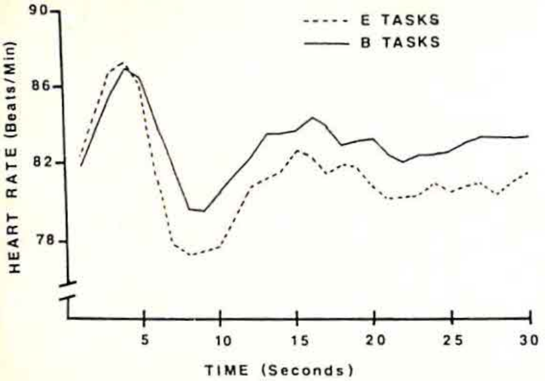


FIG. 3. Average second-by-second HR values for E and B tasks separately, Exp. II.

Sroufe, 1971) have demonstrated relationships between HR and HR variability and respiration. Thus, the purpose of Exp. II was to determine whether changes in respiratory activity accompany the changes in cardiac activity observed in Exp. I.

Several procedural changes were made in Exp. II. First, the measurement period during alert and task periods was extended and the duration of all three periods was equalized at 40 sec. This was to confirm that the HR differences between E and B tasks were persistent and to give a longer time for samples of HR following the response to the onset of the task. Second, a change in the scoring procedure for searching speed on checking tasks was made to make the procedure similar to that used on counting tasks. Finally, errors of omission were recorded in order to provide further evidence (in addition to that from searching speed) that the discriminability variable had been successfully manipulated.

Method

Subjects.—The *S*s were three female and nine male graduate and undergraduate volunteers from the University of Illinois.

Apparatus and procedure.—Recording apparatus was the same as in Exp. I, with the addition of a device to provide continuous recording of respiration. A temperature-sensitive diode was attached to the upper lip of the *S*, just beneath the nose. This diode was connected to a Beckman Model 9853A coupler on the dynagraph.

Of the eight presentation orders used in Exp. I, four were used again here. Each order was given to three *S*s. The duration of each of the three experimental periods was 40 sec. All other procedures were the same as in Exp. I, except for the addition of a response requirement on checking tasks. The *S* was required to indicate with a check mark the last line he had scanned after being told to stop at the end of the task period.

Scoring.—1. Speed and accuracy of searching. Speed of searching for counting tasks was assessed in the same way as in Exp. I. For checking tasks, the score representing the number of lines scanned was based on the line next to *S*'s check mark. Accuracy was assessed by establishing the number of possible identifications (given the number of lines scanned) and then determining the actual number of identifications either from *S*'s written total (counting tasks) or by counting check marks (checking tasks).

2. Heart rate. The computer was again used to provide second-by-second analysis of HR. For each experimental period, second-by-second HR values were obtained for 30-sec. samples. The samples were taken from the first 30 sec. of alert and task periods, and the last 30 sec. of the rest periods.

3. Respiration. For each *S*, and each of the 16 alert, rest, and task periods, a measure of respiratory period was obtained. This was defined as the time, in seconds, of the last three complete respiratory cycles occurring during the relevant period and was measured by hand scoring of the chart recording. Since the paper speed was 1 mm/sec, and measurements were made to the nearest millimeter, effective measurement of respiratory period was to the nearest second.

Results

Analysis of tonic HR, cardiac deceleration, and change in tonic HR.—Figure 3 shows the results of second-by-second analysis of HR during the task periods. As in Exp. I, mean HR values were obtained for each *S* and for each of the 16 tasks. These values represented the average HR for Sec. 16 through 30 of the task periods. The 16 values obtained were reduced to 8 by averaging across pattern, and an analysis of variance was performed on the resulting data. HR during B tasks was higher than during E tasks (see Fig. 4), $F(1, 11) = 8.83$, $p < .05$. HR was also higher during checking (82.33) than counting (81.23) tasks, $F(1, 11) = 8.65$, $p < .05$. Similar analyses were performed on mean HR values for rest and alert periods. These values were based on the last 15 sec. of the rest periods and Sec. 16 through 30 of the alert periods. None of the main effects or interactions were significant in these two analyses.

Measures of cardiac deceleration were

obtained as in Exp. I by computing the difference in mean HR between alert and task periods. There was a significant difference between E and B tasks in the change from alert to task periods (see Fig. 4), $F\ (1, 11) = 7.88$, $p < .05$. Further, HR decreased for counting conditions (−.56) but increased for checking (.92), $F\ (1, 11) = 13.82$, $p < .01$.

Tonic HR changed significantly across experimental periods, $F\ (2, 22) = 3.61$, $p < .05$. Scheffé tests indicated that the comparison between rest and alert-plus-task periods was significant (see Fig. 4).

Analysis of HR variability.—As in Exp. I, values for *SD* of HR were obtained for each *S* and each experimental period using those second-by-second HR values which had been used to compute mean HR. None of the main effects or interactions were significant in three separate analyses for rest, alert, and task periods. HR variability also did not change differentially for the different tasks from alert to task periods. However, variability did change across the three periods, $F\ (2, 22) = 28.82$, $p < .001$. Scheffé tests indicated that variability was lower in the task periods (3.49) as compared to both alert (4.81) and rest (5.22) periods.

Analysis of respiration.—Several analyses of variance, identical to those performed on mean HR data, were performed on the data for respiratory activity for the three experimental periods. During the task periods, respiratory period on E tasks was longer than that on B tasks under high-density conditions, but there was no difference between the tasks under low-density conditions, $F\ (1, 11) = 14.44$, $p < .01$. The analysis of the values for the difference in length of respiratory period between alert and task indicated that respiratory period decreased more for B tasks than E tasks (see Fig. 4), $F\ (1, 11) = 8.64$, $p < .05$. Respiratory activity also changed significantly across the three experimental periods, $F\ (2, 22) = 34.09$, $p < .001$. Respiratory period was shorter in task periods than in either rest or alert periods (Scheffé tests).

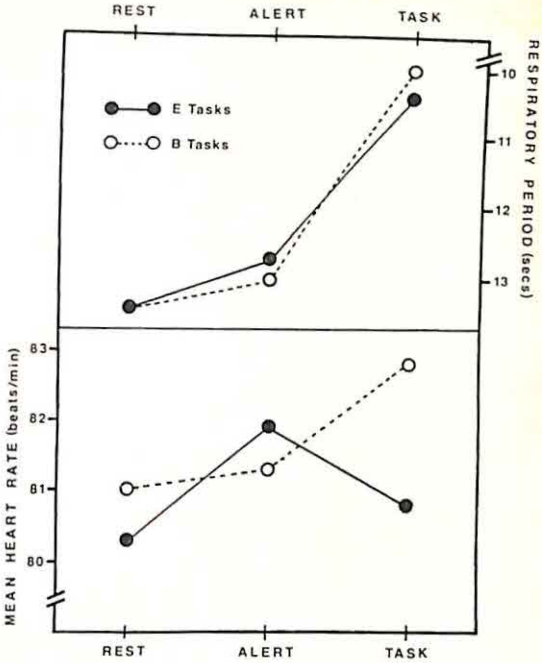


FIG. 4. The upper graph shows values for the length of respiratory period (time for three complete respiratory cycles) for the three experimental periods. (The lower graph shows mean HR values, representing tonic HR during the three periods, Exp. II.)

Analysis of searching performance.—The results for searching speed were the same as those obtained in Exp. I except that the means were higher in this experiment. A chi-square test was used to determine whether the proportion of errors of omission to possible detections was greater for E than B tasks. The *S*s missed 7% of the 710 possible identifications on E tasks, and .6% of the 1,438 possible identifications on B tasks ($\chi^2 = 68.99$, $p < .001$).

Respiration and searching speed.—As in Exp. I, the 16 tasks performed by each *S* were divided into a fast and slow group. The mean change in length of respiration period from alert to task periods was computed for each *S* for fast and slow tasks, respectively. Although the length of respiration period decreased more for slow (−2.7) than for fast (−2.5) tasks, this difference was not significant, $t = .69$, $p > .05$.

DISCUSSION AND CONCLUSIONS

The mean values for lines scanned were higher in Exp. II, reflecting the 5-sec. increase in time allowed for searching. As in Exp. I, more lines were scanned during B tasks and counting conditions than during their counterparts. The significant Target Letter × Response Mode interaction, obtained in both experiments, merely indicates that the ratio of lines scanned in E and B tasks was constant across the two response conditions.

The data for accuracy of searching confirmed that the manipulation of target-letter discriminability was successful, since the proportion of possible identifications that were missed was greater for E than for B tasks. Although the actual percentage values were small, it is possible that this difference contributed to the observed difference in HR between E and B tasks. One implication of the error difference is that fewer identifications were made during E tasks. However, this is also implied by the finding that E tasks were performed slower than B tasks. For reasons given in the discussion of Exp. I, the difference between E and B tasks in the frequency of identifications cannot be used to account for the difference in HR between the two types of task.

The results obtained for HR in Exp. II in general confirm those obtained in Exp. I. Comparison of Fig. 1 and 3 indicates that the same form of HR response to the task was obtained in both experiments. Figure 3 also shows that the differences in HR between E and B tasks were persistent and stable. Again, HR was lower in E tasks than in B tasks.

The change in tonic HR from alert to task periods was only in the direction of deceleration for E tasks in Exp. II, whereas deceleration was observed for both E and B tasks in Exp. I. Procedural differences and differences in the *S* population may account for the discrepancy. The same differences may also explain the absence, in the second experiment, of an effect of target letter on tonic HR during rest and alert periods preceding E and B tasks. However, there is no ready explanation for the failure to replicate a significant Target Letter × Density interaction in the analysis of cardiac deceleration. Whatever the explanation for these discrepancies, they do not detract from the principle findings of differential changes in HR and different levels of tonic HR for E and B tasks.

Both experiments provide evidence for a relationship between both tonic HR and cardiac deceleration and the discrimination requirements of visual search tasks. If the objective manipulation of discriminability is accepted as a manipulation of the requirements of environmental intake, then, since HR was lower and cardiac deceleration greater for low-discriminability tasks, the results support the intake-rejection hypothesis. The results also imply the rejection of the effort hypothesis because low-discriminability tasks intuitively require more "perceptual effort." It may be argued that the results also imply the rejection of a "motor effort" hypothesis. Tests of the effects on HR of frequency of identification and speed of searching, which would both seem to be related to motor effort, were not significant. However, a possible effect of motor effort was revealed in the differences in tonic HR and cardiac deceleration between counting and checking tasks.

In the experiments of Lacey and his coworkers (Lacey et al., 1963), tonic HR was found to increase significantly from rest to alert periods and then to decrease to a lower level during intake tasks. The results of the two experiments reported here failed to replicate all these findings (see Fig. 2 and 4). This may have been due to procedural differences between these experiments and those of Lacey and his co-workers. For example, in the present situations, it seems unreasonable to expect *S*s to remain alert for 40 sec. when they know that they do not have to do anything until the end of the alert period.

The length of respiratory period changed across the three experimental periods in a manner similar to HR variability (cf. Sroufe, 1971). However, unlike HR variability, the measure of respiration period also showed a differential change from alert to task for the two discriminability conditions. Thus, respiration period changed as a function of the specific attentional demands of the different tasks. This result contrasts with those of Porges and Raskin (1969), who found that respiratory activity did not change according to specific attentional requirements. Respiration period also showed a differential change for slow and fast tasks, with slow tasks showing the greatest decrease in the length of the period. Although not significant, this result shows that the differences between E and B tasks in searching speed cannot account for the differential change in respiratory period between the two types of task. B tasks, which were performed faster, were associated with greater decreases in the duration of respiratory period.

Although there is some similarity between the data for respiratory activity and those for cardiac activity, it is clearly not possible to infer from these experiments whether this similarity is due to a causal relationship between respiration and HR. However, whatever the relationship, it is clearly complex since measures of respiratory period and those of HR and HR variability only showed partially parallel trends (for example, see Fig. 4).

To summarize the results of both experiments, HR variability varied as a function of general attentional requirements, while tonic HR and change in tonic HR (cardiac deceleration) varied as a function of specific attentional requirements. Change in the length of respiration period varied as a function of both specific and general attentional requirements.

Although not designed to do so, this study has pointed to a possible relationship between HR and performance on visual search tasks. Although the result was not significant, cardiac deceleration was greater for fast, as compared to slow, tasks. If decrease in HR is related to sensory sensitivity (Lacey et al., 1963) and sensitivity influences searching speed, then this result would have been predicted. The measure of performance, namely searching speed, may not accurately reflect the sensitivity of the *S*. A method which takes account of *S*'s accuracy, such as that provided by a signal detection theory (Green & Swets, 1966), would clearly provide a better way to quantify *S*'s performance. Such a method could not be used in the present study since *S*s failed to make errors of commission.

REFERENCES

CAMPOS, J. J., & JOHNSON, H. J. The effects of verbalization instructions and visual attention on heart rate and skin conductance. *Psychophysiology*, 1966, **2**, 305–310.

CAMPOS, J. J., & JOHNSON, H. J. Affect, verbalization, and directional fractionation of autonomic responses. *Psychophysiology*, 1967, **3**, 285–290.

DUFFY, E. *Activation and behavior.* New York: Wiley, 1962.

EDWARDS, D. C., & ALSIP, J. E. Intake-rejection, verbalization, and affect: Effects on heart rate and skin conductance. *Psychophysiology*, 1969, **6**, 6–12.

ELLIOTT, R. Tonic heart rate: Experiments on the effects of collative variables lead to a hypothesis about its motivational significance. *Journal of Personality and Social Psychology*, 1969, **12**, 211–228.

GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

JOHNSON, H. J., & CAMPOS, J. J. The effect of cognitive tasks and verbalization instructions on heart rate and skin conductance. *Psychophysiology*, 1967, **4**, 143–150.

LACEY, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy.* Washington, D. C.: American Psychological Association, 1959.

LACEY, J. I. Somatic response patterning and stress: Some revisions of activation theory. In M. H. Appley & R. Trumbell (Eds.), *Psychological stress: Issues in research.* New York: Appleton-Century-Crofts, 1967.

LACEY, J. I., KAGAN, J., LACEY, B. C., & MOSS, H. A. The visceral level: Situational determinants and behavioral correlates of autonomic response patterns. In P. H. Knapp (Ed.), *Expression of the emotions in man.* New York: International Universities Press, 1963.

PORGES, S. W., & RASKIN, D. C. Respiratory and heart rate components of attention. *Journal of Experimental Psychology*, 1969, **81**, 497–503.

SCHEFFÉ, H. *The analysis of variance.* New York: Wiley, 1959.

SROUFE, L. A. Effects of depth and rate of breathing on heart rate and heart rate variability. *Psychophysiology*, 1971, **8**, 648–655.

TURSKY, B., SCHWARTZ, G. E., & CRIDER, A. Differential patterns of heart rate and skin resistance during a digit-transformation task. *Journal of Experimental Psychology*, 1970, **83**, 451–457.

(Received April 7, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 380–388

MEMORY FOR UNATTENDED INPUT [1]

JONATHAN C. DAVIS [2]
University of Toronto, Ontario, Canada

MARILYN C. SMITH
Scarborough College

A number of faults were found with procedures that directly test *Ss'* memory for unattended input. To avoid these defects, the extent of unattended-channel processing was determined indirectly by observing the effect that the presentation of the unattended had on recall of the attended. It was concluded that unattended items are excluded from primary memory since they little affected the extent or shape of the recency effects for attended items. The evidence contradicts D. A. Norman's 1968 and 1969 theoretical position that all sensory inputs, attended or not, excite their representation in primary memory.

How thoroughly does the human observer process unattended information? Broadbent's (1958) "physical" model of selective attention assumed that just one channel of information at a time is perceived and that all other channels are filtered out after only sensory analysis and storage. The present study will reconsider not only how far unattended (UA) input penetrates the perceptual system but also what level of memory such input achieves. Early evidence from dichotic listening (Moray, 1959) suggested that following shadowing of one ear, *S*s retain little from their nonshadowed ear. Later evidence (Norman, 1969; Peterson & Kroener, 1964) obtained under a more careful control of the retention interval for the nonshadowed material revealed some short-term, but little long-term, memory for nonshadowed events. The precise nature of that short-term memory is the main interest here.

Disagreeing with Broadbent's (1958) early formulation and its later refinement by Treisman (e.g., 1964, p. 217–218), Norman (1969) and Deutsch and Deutsch (1963) argued that all channels of sensory input, attended or not, reach a postperceptual level of the information-processing system. That is to say, all signals are analyzed beyond their physical properties, being to a degree semantically processed. Filtering or, rather, the bottleneck in information extraction occurs only after this substantial processing has taken place. Implicit to this view (referred to as the late-selection position) is the premise that, for some period, a multichannel representation of information exists in short-term memory. In fact, Norman (1968, p. 528; 1969, p. 92) explicitly postulated the existence of a multichannel representation of information in *primary memory* (PM). It is important to note that Norman was referring not to a sensory-store short-term memory, such as Broadbent's S system, but to that short-term memory the properties of which seem to be explainable in terms of nonsensory, psychological units or chunks.

The present study attempted to either confirm or deny the existence of UA input to PM. Much of the published evidence that previously confirmed its existence was obtained by directly testing *Ss'* memory for items arriving on a secondary channel. The *S*s were typically made to shadow a message given to one ear and then tested for retention of material given to the other ear. Peterson and Kroener (1964), Norman (1969), and Glucksburg and Cowan (1970) employed such a procedure and found substantial retention of the nonshadowed input. Memory, which was considerably less than perfect by immediate recall testing, was typically found to fade as a function of test delay.

There are problems associated with shadowing procedures that detract from their results. First, repeated testing of the nonshadowed ear makes it functionally relevant. Consequently, even building a high degree of difficulty into the shadowing task cannot guarantee a total absorption of attention. This might cause the procedure to overestimate the extent of memory for UA messages.

---

[1] This research was supported by the National Research Council of Canada through a fellowship to J. Davis and Grant APA191 to M. Smith.

[2] Requests for reprints should be sent to Jonathan C. Davis, Department of Psychology, University of Toronto, 100 St. George Street, Toronto 181, Ontario, Canada.

A second undesirable feature of shadowing paradigms might cause underestimation of UA channel retention. Shadowing may provide interference for the UA channel of a sort analogous to what has been called output interference in memory research (Tulving & Arbuckle, 1966). Briefly, output interference, to which PM items are particularly susceptible, is a reduction in the probability of recalling an item as a result of recalling other items. Nonshadowed information, if at some point perfectly represented in memory, might be poorly recalled as a result of the continuous recall (shadowing) of attended information. Even the so-called zero-delay items, which are immediately tested items from the end of the input list, are not immune to this interference, since *S*s characteristically utter a few additional words after they are signaled to stop shadowing (Norman, 1969). In light of these and other problems associated with shadowing (see Underwood & Moray, 1971, for a more complete discussion), the development of a nonshadowing procedure might be timely.

Nonshadowing dichotic listening tasks can share some of the problems of shadowing: As soon as probe, recall, detection, or recognition of the nominally UA channel is required, switching again cannot be ruled out on subsequent trials. One might, therefore, question Bryden's (1971) use of the term "unattended" for items that *S*s were requested to ignore during input but which they had to subsequently report. In that study, an attempt was made to compare and contrast memory for attended information with memory for UA information. This was done by measuring the differential effect on performance of outputting either the attended or UA items first. The problem in interpreting the results of such a study is that one cannot say whether the "unattended" items were genuinely nonattended, even though in this case they were clearly processed differently from attended ones.

Voluntary switching is best ruled out by never testing for the content of the irrelevant input. The nonshadowing experiments which follow instructed *S* to attend to one ear and that ear was tested exclusively. The *S* was given no good reason to switch and every reason not to.

Interference effects are the definitive property of primary memory (Waugh & Norman, 1965): An item's probability of staying in PM is a decreasing function of the number of subsequently presented items (regardless of presentation rate). PM appears to be a limited-capacity store, holding about three items, from which items are lost when they are displaced by new items. Thus, if words reaching a UA ear cannot be excluded from PM, they should interfere with the retention of attended words.

The experiments below sought to determine how much UA input interferes with memory for attended input. Unless a separate PM exists for each channel, UA items which cannot be completely excluded from PM should overload it, perhaps in proportion to their information content. This overload of PM should attenuate the recency effect that is observed for attended materials, making it extend over fewer serial positions. Moreover, estimates of the apparent size of PM (measured in items), the calculation of which is described below, should be correspondingly smaller than without irrelevant input.

EXPERIMENT I

The procedure employed here involved the presentation of a sequential probe for attended items. Under the hypothesis that irrelevant items might be represented in PM, a differential overloading of PM was attempted by varying the content of the irrelevant input.

Method

Subjects.—Of the 12 paid *S*s who began, 1 dropped out leaving an extra *S* in one condition.

Apparatus and procedure.—Each to-be-remembered (TBR) list of six words was presented monaurally at a rate of $1\frac{1}{2}$ words/sec, followed by a nonspeech signal (a metallic tap) that marked the end of the list and a probe word chosen from among the first five. Regardless of the materials presented to the rejected ear, the task was to give the word that had followed the probe word in the relevant ear. The first item in the list was tested by requesting *S* to "give the first item." For a given *S*, the relevant ear was kept constant both within and across sessions to allow him every advantage in learning to exclude (if possible) the irrelevant input. For 5 of the 11 *S*s the right ear was the relevant one, and for the others the left was. To guarantee ample practice in excluding the irrelevant ear, *S*s were run for a total of seven sessions, each held on a separate day with all experimental conditions given equally often within a session. The first day was considered practice; its data will not be reported.

The use of the moderately rapid presentation rate of 1½ items/sec was to help rule out *S*'s switching between ears. Of course, under many switching models (e.g., Moray, 1969), switching between the ears is still possible at this rate but surely more difficult than at a slower rate. The instructions strongly emphasized the attending to and remembering of items given to the one ear and not the other. Moreover, the procedure reinforced the credibility of these instructions by never probing with an item from the rejected ear or otherwise misleading *S* to gain additional information. In short, the ear that *S*s were requested to ignore was functionally as well as instructionally irrelevant.

The five within-*S*s experimental conditions are defined by the type of verbal and quasi-verbal material given to the rejected ear. In one condition, the control, the irrelevant channel was in fact empty, with *S*s hearing only the monaural input of the TBR list. In two conditions, the irrelevant or to-be-ignored (TBI) material was a string of words chosen from the same pool as the TBR words. In the remaining two conditions, strings of nonsense syllables (CVCs) formed the TBI list. Conditions were further partitioned according to how TBI items were synchronized with TBR items: TBI items were either *(a)* approximately *simultaneous* with TBR items or *(b)* temporally *staggered* in nonoverlapping relation to them. In all cases, six TBI items were given per list. The synchronization of items on two separate channels has been identified as an important variable in other attention research (e.g., Moray, 1960).

In the Simultaneous condition, the onset of a TBI item actually preceded the onset of its "simultaneous" TBR partner by an average of about 25 msec. (with a standard error of about 55 msec.). This mean lead by TBI items, which was estimated from a sound amplitude pen recording of 168 sampled pairs, could possibly have made TBI items more difficult to ignore than would have been the case with truly simultaneous presentations.

One of our goals was to determine whether UA materials of varying information load and meaningfulness would *differentially* affect the retention of attended information. To help *S* avoid confusion concerning the type of TBI input occurring within each successive list and to prevent *S* from having to sample TBI material to know its content, lists of a given condition were blocked in groups of 14, and *S* was always informed of the nature of the irrelevant input in a forthcoming block. Moreover, to guard against a warm-up effect in the exclusion of TBI material, the data from the first two lists of a block were discarded. In the remaining 12 lists of a block, all serial positions of the TBR lists were probed twice in a random order. Following each list a written response, for which 10 sec. were allowed, was required along with a three-level confidence rating, of which no use has been made. In every half-session there were five blocks, each representing one experimental condition; the blocks were separated by 1 min. of rest and 15 sec. of instructions, and the half-sessions were separated by 5 min. of rest. Ordering of conditions (blocks) within a half-session was counterbalanced across days.

Materials.—The six items in each TBR list were one-syllable nouns and verbs randomly chosen without replacement for lists in a given session (but with replacement between sessions). The pool of about 800 words was constructed from the Thorndike and Lorge tables, representing all objective frequencies but excluding proper nouns, onomatopoeic words, and homophones. The six items of the TBI list were either one-syllable words chosen from the above pool or easily pronounceable CVCs chosen from the Noble (1961) norms with a meaningfulness value of between about 1.4 and 2.7. None of the CVCs were words.

To facilitate the perceptual distinctiveness of the individual words, constraints were placed on the order of words within a list. No word in a TBR list was immediately preceded or followed by an item beginning with the same letter in either ear; and, in the case of simultaneous input conditions, no TBR item began with the same letter as its simultaneously presented TBI partner. In staggered conditions, each TBI item always followed the TBR item with which it was paired.

Design.—The extra *S*'s data were temporarily set aside and the data matrix collapsed across half-sessions and the two probes of a serial position within a half-session to produce a 6 × 6 × 5 × 5 × 2 between-within design: Session × Serial Position × Condition × *S*s × Ear. The *S*s were nested under the fifth factor, and the other three factors were within *S*s. A second analysis was performed by collapsing the data matrix over *S*s and over the two probes per serial position within a half-session to yield a 6 × 2 × 5 × 6 × 2 design: Session × First versus Second Half of a Session × Condition × Serial Position × Ear. This latter analysis allowed the inclusion of the extra *S*'s data and permitted the study of performance variations within a session.

Since the data were proportions, an additional analysis of variance was computed on arc-sine transformations of the data to ensure that correlations of the means with the variances were not contaminating the statistical tests. This additional analysis did not alter the pattern of results, and its details are therefore omitted.

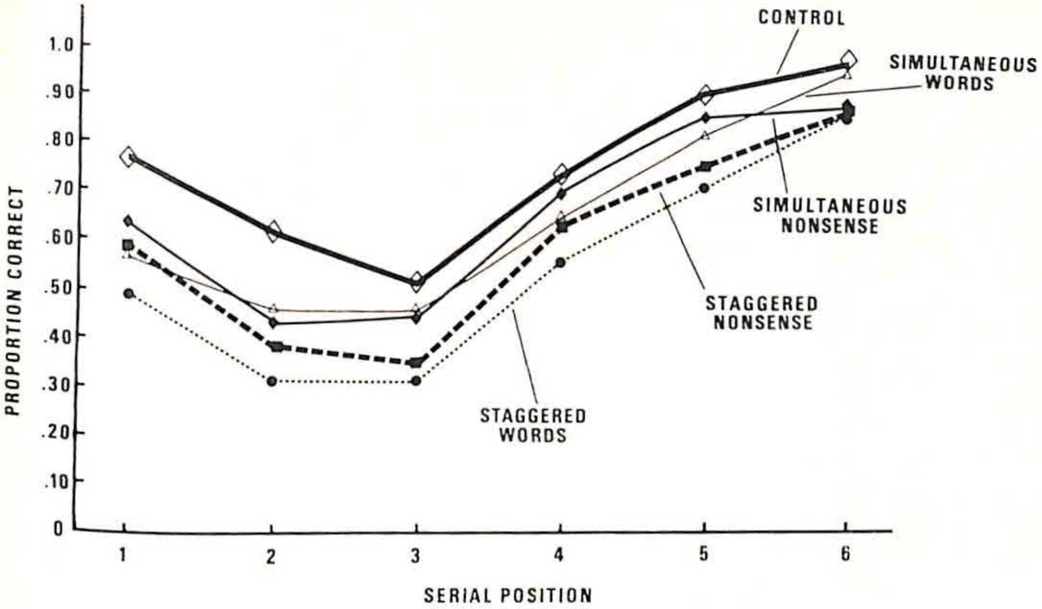


FIG. 1. Proportion of items correctly reported from each serial position, as a function of the nature and the synchronization of the irrelevant material, Exp. 1.

Results and Discussion

Figure 1 gives the serial position curves for all five conditions. As the graph suggests, serial position and condition effects were large and reliable, F (5, 545) = 227.0, and F (4, 545) = 42.1, both ps $< .001$, respectively. By far the most important result for present theoretical purposes was the finding that the recency effect was not reduced by UA input. The distinct types of UA input differentially depressed overall recall but only slightly curtailed the capacity of PM. By its customary definition, the recency effect is the higher probability of recalling those items near the end of a list. It encompasses all serial positions above the horizontal asymptote of probability, if there is one; otherwise, its extent is measured back to, but not including, the lowest point in the serial position curve. Based on the Waugh and Norman (1965) assumption of *independent* probabilities of being in primary and secondary memory, one can calculate the expected number of items held in PM by summing $\text{PM}(i) = (\text{R}(i) - sm)/(1 - sm)$ over all serial positions i in the recency effect, where $\text{R}(i)$ is the probability of recalling an item from serial position i, sm is the probability of recalling an item from the asymptote or low point in the curve, and $\text{PM}(i)$ is the probability of recalling an item in serial position i from PM. The estimated capacity of PM for the Control condition, the Simultaneous conditions, and the Staggered conditions was 2.2, 2.0, 2.0, 1.9, and 1.8 items, respectively. In the Staggered conditions, this amounted to a 14%–18% loss of capacity relative to the Control condition, whereas one might have expected more than a 50% loss in the event of unimpeded access for UA input to PM (especially since the final input item was UA). The Serial Position × Condition interaction was significant, F (20, 545) = 1.65, $p < .05$. However, this was not due to a change in the shape of the recency effect for the experimental conditions, but was due rather to a larger difference between the Experimental conditions and the Control in the first half of the list. Statistically, the Conditions × List Half interaction was significant, F (5, 545) = 3.0, $p < .05$, while the other Treatment × Serial Position contrasts were insignificant, F (15, 545) = 1.42, $p > .10$.

An orthogonal partitioning of the treatment contrasts (following the Method of Planned Comparisons, see Hays, 1963, Ch.

14) revealed that much of the variation among conditions was due to the difference between the Control condition and the others, $F(1, 545) = 107.2$, $p < .001$. No explanation of this sizable difference is offered at this point, since it bears little on the question posed by this study. Most of the remaining variation among treatments was due to synchronization; simultaneous irrelevant presentations interfered less than staggered ones, $F(1, 545) = 52.0$, $p < .001$. Concerning the remaining two degrees of treatment variation, no difference was found between irrelevant words and nonsense syllables under simultaneous presentation, $F(1, 545) < 1$. Only when the irrelevant input was staggered were words clearly more interfering than CVCs, $F(1, 545) = 10.9$, $p < .001$.

Although performance increased monotonically across sessions, $F(5, 545) = 7.0$, $p < .001$, this practice effect neither interacted with condition, $F(20, 345) < 1$, nor with the Serial Position × Condition interaction, $F(100, 445) = 1.1$. The second analysis of variance (in which *S*'s interactions comprised the error mean square) produced similar results.

Performance was better with the left ear as the irrelevant one, $F(1, 8) = 5.9$, $p < .05$. Moreover, the relatively greater difficulty in excluding irrelevant words in the staggered condition was magnified with the right ear as the irrelevant one, an effect which resulted in an Ear × Condition interaction, $F(4, 32) = 2.9$, $p < .05$, and which suggests a right ear dominance.

The cross-channel intrusion data, discussion of which is momentarily deferred, are given in Table 1.

TABLE 1

CROSS-CHANNEL INTRUSIONS AS A FUNCTION OF THEIR INPUT SERIAL POSITION, EXPERIMENT I

Input condition	Input Serial Position					
	1	2	3	4	5	6
Staggered	15	8	12	19	40	22
Simultaneous	9	7	1	16	6	4

EXPERIMENT II

The generality of the preceding results was extended by using an entirely different memory paradigm, free recall. Moreover, the range of primary memory (defined by the extent of the recency effect) was made more open to analysis by using longer lists.

Method

Subjects.—Sixteen paid undergraduates served as *Ss*.

Apparatus and procedure.—The apparatus and procedure were largely the same as before. The task was to write down in any order as many TBR words as possible from each 15-word list. The written recall (now 45 sec.) and the beginning of the next list were signaled as before. The *Ss* were told to identify a word of uncertain input (ear) origin with a check mark.

Only three conditions were used: (*a*) a control with no irrelevant input, (*b*) a staggered condition with irrelevant words, and (*c*) a staggered condition with irrelevant nonsense syllables. Staggered presentations were used and simultaneous ones omitted because in the first experiment the former magnified the potentially interesting differential interference effects of words and nonsense words. Seven lists of a given condition were blocked as before, and their order within a half-session again counterbalanced across sessions, of which there were five. The first day's data were included in the analysis so that possible interactions of the experimental effects with practice could be more closely scrutinized.

Materials.—All constraints on placement of words within a list were as before.

Design.—The data matrix was collapsed over serial positions to yield a 5 × 2 × 3 × 7 × 2 × 8 between-within design with the following factors: Session × First Half of a Session versus the Second Half × Condition × List (within a Block) × Ear × *Ss*. The *Ss* were nested under the fifth factor, which means that for a given group of eight *Ss* the ear to which TBR words were presented was always fixed. The other four factors were within-*Ss* variables. Under this analysis, performance variations within and between sessions, including the effects of extended practice, could be studied. A second analysis of variance was computed after collapsing across half-sessions, sessions, and lists within a block to produce a 3 × 15 × 2 × 7 design, representing Condition × Serial Position × Ear × *Ss*.

Results and Discussion

Although the results, which are presented in Fig. 2, were based on a free recall procedure, they largely paralleled those of the first experiment. Again there were large and reliable serial position effects, $F(14, 196) = 116.2$, $p < .001$, and effects of con-

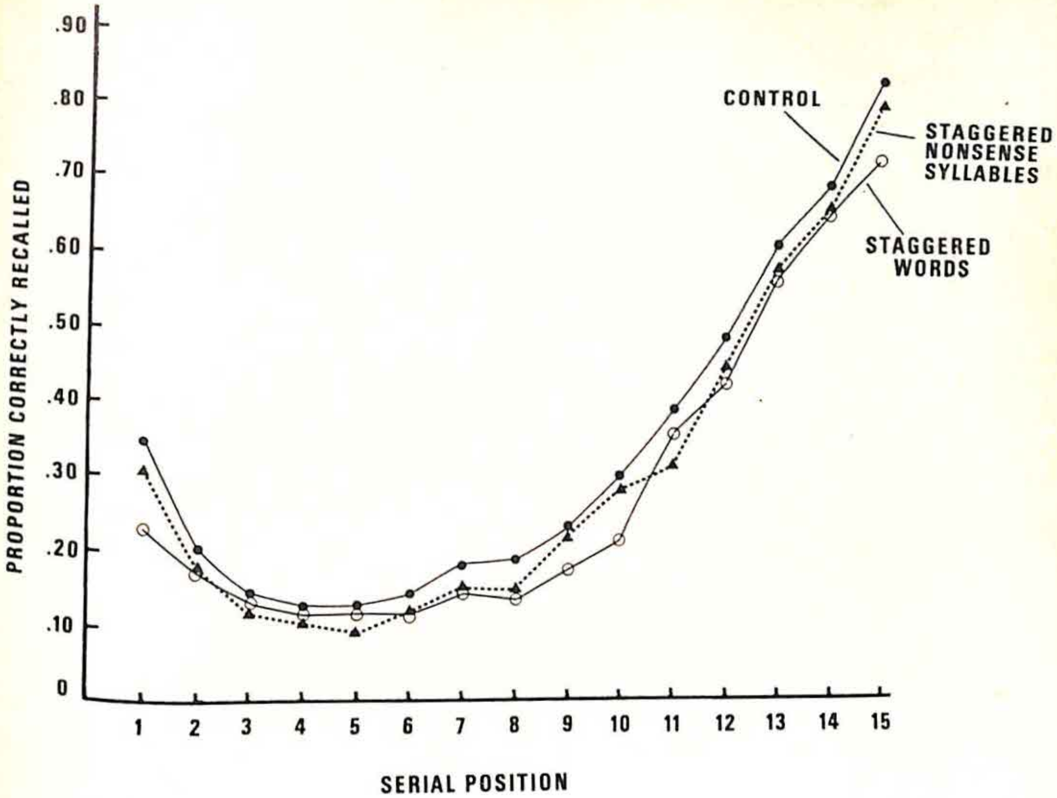


FIG. 2. Proportion of items correctly recalled from each serial position as a function of the nature of the irrelevant material, Exp. II

dition, F (2, 28) = 86.5, $p < .001$. More importantly, there were no differences among conditions with respect to the extent of the recency effect (see Fig. 2). Moreover, estimates of the size of PM (in items), calculated as in Exp. I, were practically the same for all conditions (about 2.6 items). Depending on which serial position is taken as the beginning of the asymptote, the Word condition showed from a 9% to an 11% deficit in PM capacity relative to the Control condition. Finally, unlike the first experiment, there was neither a main effect of ear nor a significant interaction between ear and condition, F (1, 14) = 1.5 and F (2, 192) < 1, respectively.

Intrusions from the UA ear (from when words were given there, of course) are plotted in Fig. 3 as a function of their input serial position. While such intrusions were relatively rare (3.2% of the recalled words), their roughly exponential decay over serial position is paradoxically suggestive of a short-term memory decay function for TBI items. The results are paradoxical because on the basis of the recall serial position curves one might like to conclude that no PM representation for TBI items exists. One obvious way out of this conceptual dilemma would be to postulate a short-term sensory storage for TBI items. Such a memory might correspond to the S system of Broadbent's (1958) original formulation, the one explicitly discounted by the late-selection modelers, including Norman (1969). Later we will consider the possibility of TBI items' residing (*a*) in a separate PM or (*b*) in neither a primary memory nor a strictly sensory store. It is this latter alternative which is ultimately favored.

The cross-channel intrusion data of the first experiment (see Table 1) did not follow the pattern of the second. Observe that the proportion of intrusions from the last TBI input serial position was smaller than the corresponding proportion in the second experiment (Fig. 3). These TBI items may have also resided in a short-term store, but

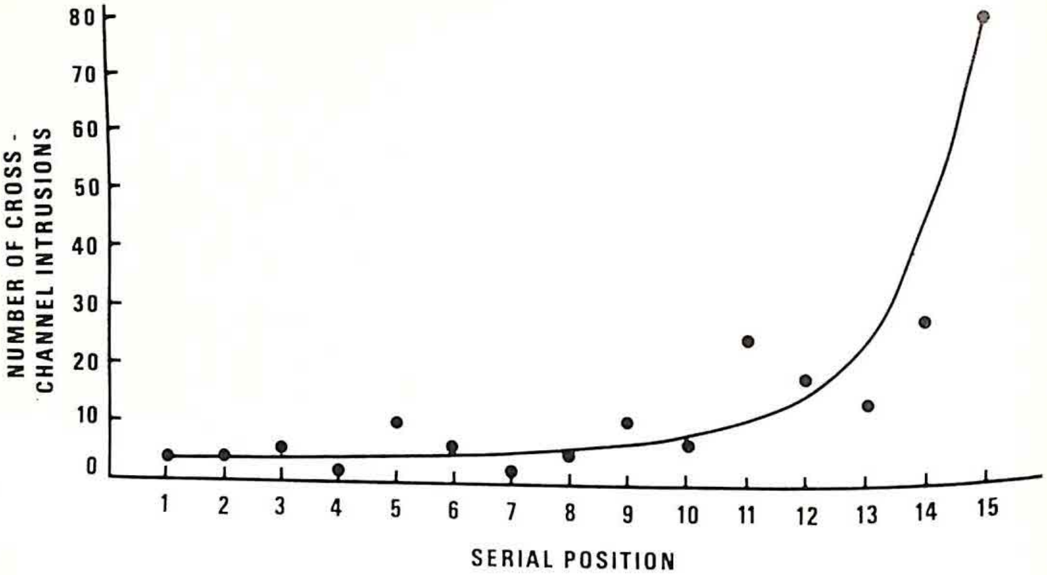


FIG. 3. Number of intrusions recalled from the irrelevant ear as a function of input serial position, Exp. II.

were so perfectly tagged by their ear of input that they were never offered as responses. Such an interpretation would make sense only if those intrusions recalled from the last TBI input position in the free recall experiment were offered after their "ear" tag had decayed, namely, in a later output serial position, i.e., after recall of other items. All intrusions from the second study, including both extralist and cross-channel ones, are given in Table 2 as a function of their mean output serial position. Note that all intrusions, including cross-channel ones, did in fact appear in later output positions. The evidence suggests, therefore, that no matter what type of store UA items were recalled from (PM or otherwise), they were recalled only after their tag of ear identity was lost.[3]

TABLE 2

MEAN OUTPUT SERIAL POSITION FOR INTRUSIONS AND CHECKED ITEMS

Unchecked cross-channel intrusions	Unchecked extralist intrusions	Checked cross-channel intrusions	Checked attended items	Checked extralist intrusions
4.18 (175)	4.57 (4047)	5.33 (46)	5.85 (355)	6.35 (1219)

Note.—The numbers in parentheses are the number of observations upon which each point is based. The mean output position for unchecked TBR items was about 2.5, since mean recall was about 5 items.

GENERAL DISCUSSION

The data clearly suggest that UA channels of information are largely excluded from PM. Of course, one might still want to argue that there are as many primary memory stores as there are input channels and that this is what accounts for the data. However, such an argument can possibly be dismissed on a priori logical grounds. If the human observer had a separate PM for each ear, for example, his recall of a list of TBR items should perhaps be enhanced by presenting the list dichotically. It is doubtful that such an enhancement would occur.

The data could perhaps be accounted for by a switching model or by an attenuation model, both of which are models traditionally favored when evidence is advanced against the late-selection position. Under these models the effective exclusion of TBI items from memory follows from observers having kept their switch (or hierarchy of switches in the case of an attenuation model) attuned to the relevant channel. The possibly greater damaging effect of staggered over simultaneous UA inputs might follow from the staggered synchronization of TBI items allowing a greater opportunity for the occasional switch. However, one implication of the present data is that such switching must occur almost automatically or involuntarily since it was not reinforced; perhaps the TBI inputs "called" the switching mechanism in a manner analogous to a peripheral device calling the control of a computer with which it is interfaced (a feature of Moray's 1969 model). A switching model might also explain the apparent drop-off in cross-channel intrusions as a function of input serial position (Fig. 3) as loss from memory of those rare items picked up during an (involuntary) switch. However, there is one possible problem with this latter interpretation. If a word were picked up during a switch to a UA channel, the word should presumably be treated thereafter as a TBR item. Assuming that such switches occur uniformly across serial positions, then the serial position curve for intrusions should mirror the recall curve, and in this case it does not. (Compare Fig. 2 and 3.) The intrusion curve lacks a primacy effect.

---

[3] Check marks, which, when used, designated words of uncertain origin, were also identified with later output serial positions. A separate analysis of the checked cross-channel intrusions was performed to see whether they were comprised largely of items from the last input serial position. They were not; they were distributed across input serial positions in the same proportions as unchecked intrusions.

The abundance of evidence for parallel semantic analysis of both UA and unattended channels might compel one to seek a nonswitching model to account for the present data. Consider some of the following findings with which our data must be reconciled: Brown (1970) found no evidence for attenuation of input given on an irrelevant channel. His *Ss* were dichotically presented a pair of words embedded in noise. Sometimes *Ss* were informed in advance of presentation which ear they would be tested on; sometimes they were not. No difference was found in recognition of the target word as a function of *Ss'* prior knowledge of the relevant ear. Lewis (1970) showed that the reaction time to shadow words given to one ear was affected by their semantic relationship with irrelevant words concomitantly heard in the other ear. McKay (1971), for another, found that the semantic interpretation of an attended, but ambiguous, sentence given to one ear could be biased by a TBI cue word given to the other ear. For example, if the sentence "The boy threw the rock against the bank" were given to one ear, and if either the word "money" or "river" were presented to the other ear, the ensuing interpretation of "bank" could be biased in the obvious fashion. Clearly these findings suggest some automatic semantic analysis of UA input in parallel with attended-channel processing. (Of course, McKay's results might be accounted for by a switching model in which the mechanism that runs the switch performs analysis-by-synthesis (Neisser, 1967). Under such a model, a sample of other channels might be taken whenever ambiguities arise in the interpretation of what is heard on the "attended" channel.)

Suppose then that all input channels are routinely given some semantic analysis. How then might one account for the present data? The conceptualization of processes favored here is borrowed not from the attention literature but from memory work. The main implication of the data is that although TBI items may be analyzed for meaning, it is at little or no cost to the short-term retention of TBR items. The situation may therefore be analogous, perhaps identical, to an intentional forgetting paradigm (Bjork, 1970; Davis & Okada, 1971). The *Ss* may be momentarily aware or conscious of each TBI event (including its "meaning") but may be able to quickly and efficiently forget it. What corresponds to the cue to forget, the physical tag of ear origin, may be apprehended simultaneously with (or immediately after) perception.

What might the present data specifically imply for a better understanding of memory in attention (and perhaps attention in memory)? Short-term memory might now be considered item- or chunk-sensitive only with respect to the retention of TBR items. If PM is that short-term store of three or four items from which information is lost when new items displace old, then PM clearly contains only attended items. If, on the other hand, PM is considered to be a stage of processing (rather than a "store"), then its defining interference properties are now known to be derived solely from the processing interference of TBR items. The probability of recall of a TBR item did not diminish as a function of the number of subsequent TBI items, only as a function of the number of subsequent TBR items. The basis for this claim is the finding that UA input had little effect on the shape or extent of the recency effect for attended items.

Unless the data reported by others suggesting parallel semantic analysis of all sensory channels is discounted, it might be useful to consider "unattended" input as information irrelevant to the system, partially comprehended perhaps, but then efficiently forgotten. In any case, our main conclusion is clear: Primary memory need only be for pertinent items, instead of for all incoming items. In short, we have demonstrated that selection can be early, before PM.

REFERENCES

Bjork, R. A. Positive forgetting: The non-interference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 255–268.

Broadbent, D. E. *Perception and communication.* New York: Pergamon, 1958.

Brown, J. Recognition and the direction of attention. In A. F. Sanders, (Ed.), *Attention and performance III.* Amsterdam: North Holland Publishing Company, 1970.

Bryden, M. P. Attentional strategies and short-term memory in dichotic listening. *Cognitive Psychology*, 1971, **2**, 99–116.

Davis, J. C., & Okada, R. Recognition and recall of positively forgotten items. *Journal of Experimental Psychology*, 1971, **89**, 181–186.

Deutsch, J. A., & Deutsch, D. Attention: some theoretical considerations. *Psychological Review*, 1963, **70**, 80–90.

Glucksburg, S., & Cowan, G. N. Memory for nonattended auditory material. *Cognitive Psychology*, 1970, **1**, 149–156.

Hays, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

Lewis, J. L. Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology*, 1970, **85**, 225–228.

McKay, D. G. Aspects of a theory of comprehension and attention. Unpublished study, University of California, Los Angeles, 1971.

Moray, N. Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 1959, **11**, 56–60.

Moray, N. Broadbent's filter theory: Postulate H and the problem of switching time. *Quarterly Journal of Experimental Psychology*, 1960, **12**, 214–220.

Moray, N. *Attention: Selective processes in vision and hearing.* New York: Hutchinson Educational, 1969.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Noble, C. E. Measurements of association value, rated associations, and scaled meaningfulness for the 2100 CVC combinations of English alphabet. *Psychological Reports*, 1961, **8**, 487–521.

Norman, D. A. Toward a theory of memory and attention. *Psychological Review*, 1968, **75**, 522–536.

Norman, D. A. Memory while shadowing. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 85–93.

Peterson, L. R., & Kroener, S. Dichotic stimulation and retention. *Journal of Experimental Psychology*, 1964, **68**, 125–130.

Treisman, A. Verbal cues, language, and meaning in selective attention. *American Journal of Psychology*, 1964, **77**, 206–219.

Tulving, E., & Arbuckle, T. Y. Input and output interference in short-term associative memory. *Journal of Experimental Psychology*, 1966, **72**, 145–150.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

Underwood, G., & Moray, N. Shadowing and monitoring for selective attention. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 284–295.

(Received April 10, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 389–395

IRRELEVANT INFORMATION AND PROCESSING MODE IN SPEEDED DISCRIMINATION

HAROLD L. HAWKINS[1] AND R. HAL SHIGLEY

University of South Florida

The *S*s were presented with stimuli of varying complexity which they were required to categorize rapidly as either SAME or DIFFERENT. When stimuli contained no irrelevant features or when they contained irrelevant features that were compatible in state (SAME or DIFFERENT) with those defined as relevant, reaction time (RT) was not affected by stimulus complexity. However, when stimuli contained irrelevant features incompatible in state with those defined as relevant, RT systematically and markedly increased with the number of relevant features. This latter finding suggested the presence of a limited capacity mechanism, operating in the serial mode possibly, which functions only when stimuli contain irrelevant-incompatible information. Evidence was presented which indicates this mechanism provides a match–mismatch comparison between the features of stimulus input and a stored list of stimulus features previously defined by instruction as relevant. Such a comparison is necessary, it was argued, to insure that responses are based solely on the relevant features of stimulus input.

Studies investigating the mode of information processing in speeded visual discrimination have yielded conflicting results. In the speeded discrimination task, *S* is presented on each trial with a pair of multidimensional test figures which must be categorized quickly and accurately as either physically identical (SAME) along all relevant dimensions or DIFFERENT along at least one relevant dimension. In a study of performance in this task, Egeth (1966) has concluded that the SAME–DIFFERENT judgment is based on an analytic, dimension-by-dimension interrogation of test stimuli. He concluded further that the mode by which the dimensions of complex stimuli are compared is serial; i.e., the constituent features of test stimuli are examined one after the other in a recursive fashion. Based on a study similar in many respects to Egeth's, Hawkins (1969) also concluded that complex visual stimuli are treated analytically in the speeded discrimination situation. However, contrary to Egeth's findings, the Hawkins data seemed best described by a model featuring the assumptions that stimulus dimensions are interrogated simultaneously, in the parallel mode, and that the time needed to reach a decision regarding the state of a particular dimension varies from trial to trial (distributed decision times).

A number of procedural differences exist between the Egeth (1966) and Hawkins (1969) studies. The purpose of the present experiment was to examine the possibility that one of these differences—namely, that Egeth's stimuli contained irrelevant information under most conditions whereas those used by Hawkins did not—produced the difference in processing mode apparent between the two studies. The selection of this particular variable for study in the present experiment was prompted in part by an analysis of Egeth's SAME RT data. Serial processing models imply that SAME RT should increase systematically with the number of stimulus dimensions which must be discriminated. Depending on whether decision times are treated as a constant or as a random variable, parallel models imply either no SAME RT increase with increasing stimulus dimensionality, or an increase smaller than that implied by serial models. To comparatively evaluate these models, Egeth observed SAME–DIFFERENT RT across three levels of stimulus complexity, varied instructionally. The physical dimensionality of stimuli remained constant across levels of complexity. However, under unidimensional conditions, *S*s were told prior to each trial block to attend to only one of the three dimensions present in stimuli, and under bidimensional conditions, to attend to only two of the three dimensions present. Thus, stimuli contained irrelevant information under uni- and bidimensional but not under tridimensional conditions. Under these circumstances, SAME RT increased markedly between uni- and bidimensional conditions and, surprisingly, decreased markedly between bi- and tridimensional conditions. Consistent with the interpretation offered earlier in this article, the magnitude of the SAME RT increase between uni- and bidimensional conditions, both containing irrelevant information in the Egeth study, was well beyond that expected on the basis of any of the parallel models considered by Egeth (Hawkins, 1969). The decline in SAME RT between bi- and tridimensional conditions is also consistent with the view offered earlier in this article which implies that stimulus dimensions should be processed sequentially under the former and in parallel under the latter condition.

---

[1] Requests for reprints should be sent to Harold L. Hawkins, Department of Psychology, University of South Florida, Tampa, Florida 33620.

While the foregoing evidence certainly buttresses the view that processing mode in speed discrimination depends on whether comparison stimuli contain irrelevant features, it is no more than suggestive. The present experiment represented an attempt to provide a stronger and more systematic test of this notion.

METHOD

Subjects.—The *S*s were six undergraduate students majoring in psychology at the University of South Florida. Course credit was given for participation.

Incentive system.—To encourage *S*s to respond both quickly and accurately, a $5.00 bonus was awarded to the *S* exhibiting the lowest average "error-corrected" $\overline{\text{RT}}$ across all conditions of the experiment. The error correction consisted of adding 5 msec. to the mean RT ($\overline{\text{RT}}$) under a condition for each incorrect response made under that condition.

Apparatus.—Stimuli were viewed through a Polymetric two-field mirror tachistoscope at a distance of 43.18 cm. On a desk in front of *S* were two response keys labeled SAME and DIFFERENT, respectively. Response latency was indicated by a Hunter Klockounter. During the experiment *S* wore earphones over which a low-amplitude masking noise was channeled.

Stimuli. The stimulus figures were geometric forms (Hawkins, 1969) constructed of colored cardboard and mounted on white index cards. The horizontally arrayed stimulus pair appearing on each trial subtended a visual angle of 6° on the horizontal and 2.4° on the vertical plane.

During each trial block, unidimensional stimuli were judged either on the basis of color, form, or size. The color (C) deck consisted of pairs of 16-mm.-sq. figures, either blue or green in hue. The two hues were approximately equal in saturation and brightness. The form (F) deck consisted of pairs of neutral gray trapezoids (approximately equal in brightness to the hues used in C), which were 16 (height) × 15 (short base) × 17 (long base). The long base was either up or down. The size (S) deck consisted of pairs of neutral gray 16-mm. squares each containing an inked circle (No. 3 Rapidograph) either 7 or 9 mm. in diameter.

The bi- and tridimensional stimuli contained the appropriate combination of features described for the unidimensional stimuli. The three bidimensional decks were color/form (CF), color/size (CS), and form/size (FS). Cards comprising the CF deck consisted of pairs of trapezoids (long base up or down) either blue or green in hue. The CS cards consisted of pairs of squares (blue or green) containing inked circles that were either 7 or 9 mm. in diameter. The FS cards consisted of pairs of neutral gray trapezoids containing inked circles that were either 7 or 9 mm. in diameter. The color/form/size (CFS) cards consisted of pairs of trapezoids (long base either up or down) either blue or green in hue, and each contained an inked circle either 7 or 9 mm. in diameter.

Only one-fourth of the cards contained stimuli that were SAME so as to eliminate the contingent probabilities appearing in stimulus decks of the type used here when SAME and DIFFERENT pairs are equated in probability of occurrence (Hawkins, 1969).

Condition NI (no irrelevant information).—Three 32-card decks, C, F, and S, were used under the unidimensional conditions. Stimulus pairs were identical on 8 (one-quarter) of the cards in each of these decks and different on 24.

Three 32-card decks, CF, CS, and FS, were used under the bidimensional conditions. Each deck contained 24 cards on which stimuli differed. On 8 of these, differences appeared along one dimension; on another 8, differences appeared along the other; and on the remaining 8, differences occurred along both dimensions.

A single deck of 112 cards was used under the tridimensional condition (CFS). Prior to presentation of this deck, the entire deck was shuffled and the first 96 cards from the deck were used. Twenty-eight (one-quarter) of the cards contained stimuli identical along all three dimensions. The DIFFERENT cards consisted of seven equal-sized sets of 12 cards each. The dimensions along which they differed were C, F, S, CF, FS, and CFS. The tridimensional deck was included so that *S*s would receive the same total number of trials under Cond. NI as under Cond. I.

Condition I (irrelevant information). Six 32-card decks were used under the unidimensional conditions. The decks were C(F), C(S), F(S), F(C), S(F), and S(C) where the letter in parentheses denotes the dimension that was to be ignored by *S*. Again the stimulus pairs were identical along the relevant dimension on only one-quarter of the cards in each of these decks. On one-half of both the SAME and DIFFERENT pairs, stimuli were identical along the irrelevant dimension.

Three 32-card decks were used under the bidimensional conditions. The decks were CF(S), CS(F), and FS(C). Each deck contained 24 cards on which stimuli differed. On 8, differences appeared along one dimension; on 8, differences appeared along the other; and on the remaining 8, both dimensions differed. Again, on one-half of all trials, in each of the contingencies described stimuli differed along the irrelevant dimension.

A stimulus pair presented within Cond. I was assigned to one of two subconditions based on the relationship within that pair between the states of relevant and irrelevant dimensions. If the states of relevant and irrelevant features agreed—e.g., both were SAME or both were DIFFERENT—the pair was assigned to the Irrelevant-Compatible condition (I-C). If the states of relevant and irrelevant features disagreed—e.g., stimuli were SAME along their relevant dimensions and DIFFERENT along their irrelevant dimension, or vice versa—the stimulus pair was assigned to the Irrelevant-Incompatible condition (I-I).

Procedure.—Each of the six *Ss* participated in all treatment conditions of the experiment. Half of the *Ss* were instructed to respond SAME with their dominant hand and DIFFERENT with their other hand, while the other half were given the opposite response assignment. All *Ss* served in six experimental sessions—three in which there was no irrelevant information (NI) and three in which there was irrelevant information (I).

Within Cond. NI, three separate orders of presentation of the seven conditions were constructed for each *S*. Each order contained the following features: (*a*) The three unidimensional conditions followed one another in a block, as did the three bidimensional conditions, (*b*) the order of the uni-, bi-, and unidimensional blocks formed a 3 × 3 Latin square over the 3 days of practice; and (*c*) within the uni- and bidimensional blocks the order of particular conditions formed a Latin square across days.

Each *S* received 32 trials in each of the uni- and bidimensional conditions and 96 trials in the tridimensional condition (CFS).

Within Cond. I, three different orders of the nine conditions were constructed for each *S*. C(F), F(S), and S(C) constituted one block; C(S), F(C), and S(F) constituted the second block; and CF(S), CS(F), and FS(C) made up the bidimensional block. Each of the three orders had the following features: (*a*) The unidimensional conditions followed each other within their blocks, as did the three bidimensional conditions; (*b*) the order of the two uni- and one bidimensional blocks formed a 3 × 3 Latin square over the 3 days of trials; and (*c*) within the uni- and the bidimensional blocks the order of particular conditions formed a Latin square over the 3 days of practice. Each *S* received 32 trials in each of the uni- and bidimensional conditions. The order in which *Ss* received (I) and (NI) trials was balanced over the 6 days of practice.

Before each condition, *Ss* were told which dimensions were relevant. There were short rest breaks between the blocks of uni-, bi-, and tridimensional trials.

Before the first experimental session, a practice session was given which consisted of a series of 20 SAME–DIFFERENT trials judging letter pairs (AA, AB, BA, BB). The practice served to insure that *Ss* were familiar with their respective response assignment.

RESULTS

Same reaction time data.—The SAME RT data were broken down according to stimulus condition (C, F, S, CF, and FS), level of practice (Days 1, 2, and 3), and type of irrelevancy—no irrelevant information (NI), irrelevant-compatible information (I-C), and irrelevant-incompatible information (I-I).

The median of each *S*'s latencies under each condition was taken and the SAME RT value appearing in Table 1 represents the mean, across *Ss*, of these medians. A minor variant of this procedure was used under Cond. I-C and I-I, in which each dimension, when relevant at the unidimensional level, was represented by two decks, one with each of the remaining dimensions irrelevant; e.g., the color dimension was represented by decks C(F) and C(S). In these cases, *Ss* median was taken for each deck, the mean of the two medians was computed and the value appearing in the table represents the mean, across *Ss*, of these means.

A major objective of this experiment was to determine whether or not the presence of irrelevant stimulus information influences processing mode in speeded discrimination. Egeth (1966) has described several specific parallel-processing models. Among these, the one which has gained the most empirical support (Hawkins, 1969) is one including the assumption that

TABLE 1

"SAME" REACTION TIME (IN MSEC.) AND PERCENT INCORRECT RESPONSES (IN PARENTHESES) AS A FUNCTION OF CONDITION OF IRRELEVANCY, STIMULUS CONDITION, AND LEVEL OF PRACTICE

Day	Measure	No irrelevant information						Irrelevant-compatible information						Irrelevant-incompatible information					
		C	F	S	CF	CS	FS	C	F	S	CF	CS	FS	C	F	S	CF	CS	FS
1	RT	368	490	547	484	550	570	369	494	550	502	554	579	408	544	585	692	685	740
	Errors	(3.2)	(5.0)	(5.0)	(4.8)	(6.1)	(7.9)	(4.0)	(4.7)	(6.1)	(5.3)	(3.9)	(7.5)	(5.3)	(6.2)	(5.9)	(9.1)	(8.5)	(9.9)
	PUB				498	555	565				510	562	564				555	599	604
2	RT	364	448	507	451	502	503	355	439	493	448	494	512	386	480	538	560	607	599
	Errors	(1.8)	(4.0)	(7.1)	(4.2)	(5.9)	(7.0)	(2.8)	(4.7)	(6.0)	(4.6)	(3.2)	(7.2)	(1.9)	(6.0)	(5.3)	(8.8)	(8.0)	(7.5)
	PUB				460	517	521				450	505	505				490	550	554
3	RT	366	432	485	437	481	475	370	432	457	435	466	459	391	472	519	509	536	554
	Errors	(2.8)	(3.2)	(4.1)	(4.2)	(5.7)	(7.5)	(3.1)	(4.0)	(3.9)	(4.7)	(5.0)	(6.0)	(2.6)	(6.2)	(8.1)	(4.9)	(7.4)	(7.7)
	PUB				443	495	499				439	467	468				484	529	531

Note.—RT = reaction time; PUB = parallel upper bound estimates; C = color; F = form; S = size; CF = color/form; CS = color/size; FS = form/size. Also included under bidimensional conditions are parallel upper bound estimates generated by means of a formal procedure described by H. L. Hawkins (1969).

the time required to make a decision regarding the state of a particular dimension varies randomly from trial to trial (distributed decision times).

This model implies SAME RT will increase with stimulus complexity, but the magnitude of the predicted increase is less than implied by serial processing models. To sharpen the difference between the parallel/distributed and serial models, Hawkins (1969) devised a formal procedure for estimating the upper limit based on the parallel/distributed model of the increase in SAME RT between any unidimensional condition, Y, and a bidimensional condition, XY, where RT to Y at the unidimensional level is greater than to X. This procedure was applied to the data of the present experiment. The resulting parallel upper bounds (PUBs) appear in Table 1 in conjunction with the observed latency under each bidimensional condition for each day of practice. Observed values greater than the associated PUB were interpreted as reflecting some kind of underlying serial process.

Under Cond. NI only one of the nine obtained latencies for bidimensional stimulus pairs exceeded the associated PUB. The single exception occurred on the first day of practice for the most difficult bidimensional deck, FS. Under Cond. I-C two of the nine bidimensional latencies, FS on Days 1 and 2, exceeded PUBs. However, when stimuli contained an incompatible irrelevant dimension (Cond. I-I) the observed bidimensional RTs considerably exceeded the PUB in all nine possible cases.

The data appearing in Table 1 were subjected to an analysis of variance in which stimulus condition, condition of irrelevancy, and day of practice were all treated as repeated-measurement variables. This analysis revealed that condition of irrelevancy (A), $F(2, 10) = 24.2$, stimulus condition (B), $F(5, 25) = 31.9$, day of practice (C), $F(2, 10) = 14.6$, and the BC interaction, $F(10, 50) = 3.3$, were all significant at beyond the .01 level. The AC interaction, $F(4, 20) = 4.0$, and the

ABC interaction, $F(20, 100) = 2.0$, were significant at beyond the .05 level.

To gain more detail regarding the effects of condition of irrelevancy, subsequent comparisons were made using a two-tailed Fisher's least significant difference (LSD) test, with the criterion of significance set at .05. This analysis showed that the main effect of irrelevancy condition was due to heightened latencies on trials where stimuli contained irrelevant-incompatible information (LSD = 22.6 msec.). No latency differences appeared between Cond. NI and I-C.

Taken together the foregoing analyses indicate that irrelevant stimulus information influenced SAME response latency in the present experiment only when the irrelevant feature was incompatible in state to the relevant features of test stimuli. The results also revealed that the effects on SAME RT of irrelevant-incompatible information were significantly greater for the bi- than for the unidimensional stimuli used in this study. Finally the significant second-order interaction obtained in the analysis of the SAME RT data revealed that the effect of irrelevant-incompatible information, especially under bidimensional conditions, declined across days of practice.

DIFFERENT RT.—The DIFFERENT RT data were also subjected to an analysis in which type of irrelevancy, stimulus condition, and level of practice were all treated as repeated measurement variables. In this case, however, Cond. I-C denotes those trials on which comparison stimuli differed along all relevant dimensions and also along the irrelevant dimension. Cond. I-I refers to trials where stimuli differed along all relevant dimensions and were identical along the irrelevant dimension. The analysis of variance revealed that condition of irrelevancy, $F(2, 10) = 25.8$, stimulus condition, $F(5, 25) = 25.54$, level of practice, $F(2, 10) = 10.01$, and the Irrelevancy Condition × Practice interaction, $F(4, 20) = 4.52$, all constituted significant sources of variance ($p < .01$). An LSD test, performed on condition of irrelevancy, showed that here again irrelevant stimulus information had an effect only when incompatible

TABLE 2

"DIFFERENT" REACTION TIME (IN MSEC.) AND PERCENT INCORRECT RESPONSES (IN PARENTHESES) AS A FUNCTION OF CONDITION OF IRRELEVANCY, STIMULUS CONDITION, AND LEVEL OF PRACTICE

Day	Measure	No irrelevant information						Irrelevant-compatible information						Irrelevant-incompatible information					
		C	F	S	CF	CS	FS	C	F	S	CF	CS	FS	C	F	S	CF	CS	FS
1	RT	386	448	501	444	431	544	372	454	462	504	462	558	403	482	532	545	542	631
	Errors	(1.7)	(4.3)	(5.0)	(3.3)	(3.3)	(4.0)	(2.0)	(5.0)	(4.1)	(3.8)	(4.0)	(3.6)	(5.2)	(6.1)	(9.2)	(3.4)	(5.2)	(6.8)
2	RT	365	417	474	398	416	475	350	415	457	420	428	490	380	441	470	472	447	544
	Errors	(1.7)	(3.0)	(3.8)	(2.8)	(4.1)	(5.6)	(1.7)	(4.1)	(5.1)	(2.9)	(3.4)	(3.9)	(3.4)	(4.4)	(5.0)	(3.1)	(4.4)	(5.3)
3	RT	357	404	443	406	435	435	362	409	455	407	403	447	402	438	484	427	435	496
	Errors	(3.1)	(3.1)	(6.0)	(3.3)	(3.8)	(5.1)	(3.0)	(4.7)	(5.0)	(3.6)	(1.9)	(4.5)	(3.6)	(4.4)	(6.1)	(3.8)	(3.8)	(5.1)

Note.—RT = reaction time; C = color; F = form; S = size; CF = color/form; CS = color/size; FS = form/size.

with the state of the relevant features of test stimuli (LSD = 28.8 msec.). The interaction of condition of irrelevancy and level of practice indicates the effect of irrelevant-incompatible information declined with practice.

DISCUSSION

This experiment was designed as a test of the idea that processing mode in speeded physical discrimination depends upon whether stimuli contain irrelevant information: If stimuli are free of irrelevancy their relevant features will be compared in parallel; if not, the comparison process operates in seriatim. The intent of this hypothesis was that the processing mode in situations of the present type presumably is determined prior to stimulus onset by the instructions which specify whether or not the oncoming stimulus pair would contain irrelevant information. A strong implication of this view is that serial processing should have been evidenced on all trials where stimuli contained irrelevancy. The data showed clearly that this is not the case: Among the bidimensional stimuli containing irrelevant information, only those exhibiting state incompatibility between their relevant and their irrelevant features were processed serially. If the states of relevant and irrelevant features corresponded, or were compatible, then the relevant features appeared to be processed in the parallel mode. This finding is of particular interest in view of the fact that *S*s could not have known prior to stimulus onset under the irrelevant information condition whether the irrelevant features of the upcoming stimuli would be compatible or incompatible with their relevant features. Taken together these considerations suggest two things about the processes underlying performance in the present task. The first is that the SAME–DIFFERENT stimulus comparison functioned as a parallel operation under all conditions of this study. The second is that when the stimuli contained irrelevancy and the result of this initial parallel comparison was state incompatibility, a subsequent operation, apparently serial in nature, was evoked.

The nature of this subsequent process was brought into focus by an examination of the differences between the experimental circumstance where the process operated (Cond. I-I) and those where it did not (Cond. NI and Cond. I-C). When stimuli contained no irrelevant information or when they contained relevant information and their irrelevant dimensions were compatible in state with those that were relevant, there was no logical necessity for the relevancy (or irrelevancy) of the constituent features of the test stimuli to be ascertained. This is obviously true in the first of these two cases, where stimuli contained no irrelevant features and *S* was informed of this prior to stimulus onset. It is no less true in the second case inasmuch as the correctness of *S*s decision regarding the overall state of test stimuli logically was independent of the relevancy of any feature or combination of features present in the stimuli; i.e., all dimensions present in test stimuli unanimously indicated a single decision—either SAME or DIFFERENT.

The *S*'s task under Cond. I-I was a more complex one in that the relevant and irrelevant features of test stimuli offered incompatible information regarding the overall state of the stimuli pair. It, therefore, was essential in this case that *S* distinguish between those features of the stimuli which were relevant and those which were not. We are inclined toward the view that it was this process, of distinguishing the relevancy of the various features of the test stimuli, that constituted the apparently serial component of the latencies obtained under Cond. I-I. In developing the logic behind this position we wish to put forward several initial assumptions. First, we assume that the physical SAME–DIFFERENT comparison of simultaneously presented stimuli is a parallel process occurring at a relatively early stage in the information-processing sequence, in advance of the point at which input makes contact with memory (cf. Neisser & Beller, 1965). Second, we assume that those aspects of test stimuli defined in the experimental context as relevant are represented in memory as a list to which the components of stimulus input must be compared in order to ascertain relevancy. Third, we assume that the comparison between input and the stored array of relevant features is a serial operation (Sternberg, 1966) which, with extended training, becomes increasingly parallel (Neisser, Novick, & Lazar, 1963; Wattenbarger, 1968). Considered together, these assumptions offer the following picture of the processes underlying performance in the present task. Upon stimulus input, and regardless of whether or not stimuli are free of irrelevancy, representations of the various properties of stimulus input are compared in parallel. No consideration is given to relevancy at this point. If stimuli contain no irrelevant information, or if they do, but all dimensions are judged to be in state agreement, the output of this parallel comparison is used as the immediate basis for response (SAME or DIFFERENT) selection. However, if stimuli are known to contain irrelevant information and if state incompatibility results from the initial SAME–DIFFERENT comparison, then memory is consulted in a seemingly serial fashion to determine which of the state-labeled features comprising stimulus input are relevant. Based on the available evidence concerning the effects of practice on the mode of memory scan (e.g., Neisser et al., 1963), we interpret the declining effect of irrelevant-incompatible information in the present study as the product of an increasing ability with practice to make the input-memory comparison in parallel.

In the foregoing analysis we operationalized the serial processing notion in terms of the observation that response latency under Cond. I-I increased (beyond a parallel upper bound) with the number of stimulus dimensions which had to be categorized as either relevant or irrelevant. Townsend (1971) has described an analysis of the theoretical distinction between parallel and serial processing which shows that latency increases beyond the parallel upper bound do not necessarily imply that processing is in seriatim. A parallel operation, functioning within a limited capacity processing system, could also yield these results. More specifically, it is possible that categorization proceeds in parallel, but the *rate* at which attributes are processed declines with increases in the number of stimulus dimensions requiring categorization. Thus the greater the number of attributes processed, the greater the total time required to process all dimensions. Based on Townsend's analysis, the data of the present experiment (like those of, e.g., Sternberg, 1966) suggest that memory scan is either a serial operation or a parallel operation the pace of which slows with increases in the complexity of the array scanned. No stronger conclusion is permissible. In either case, however, the present results seem to provide a reasonable interpretation of the discrepancy apparent between Egeth (1966) and Hawkins (1969).

REFERENCES

EGETH, H. Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, 1966, **1**, 245–252.

HAWKINS, H. L. Parallel processing in complex visual discrimination. *Perception & Psychophysics*, 1969, **5**, 56–64.

NEISSER, U., & BELLER, H. K. Searching through word lists. *British Journal of Psychology*, 1965, **56**, 349–358.

NEISSER, U., NOVICK, R., & LAZAR, R. Searching for ten targets simultaneously. *Perceptual Motor Skills*, 1963, **17**, 955–961.

STERNBERG, S. High speed scanning in human memory. *Science*, 1966, **153**, 652–654.

TOWNSEND, J. T. A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, 1971, **10**, 161–163.

WATTENBARGER, B. L. Speed and accuracy set in visual search performance. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, April 1968.

(Received April 13, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 396–399

READING SPATIALLY TRANSFORMED DIGITS[1]

RICHARD L. TAYLOR[2]

Memorial University of Newfoundland, St. John's, Newfoundland, Canada

In a YES/NO reaction time task, *Ss* decided whether words described the same number as digits. The digits appeared either normally or after 180° rotation around their X, Y, or Z axis, and were read either from left to right or from right to left. The *Ss* were fastest in reading digits rotated around their Z axis, and it was argued therefore that these flat stimuli may be processed as two or three dimensional objects in depth. Spatially transformed digits shown to the left of *Ss*' point of fixation, and read from right to left, were recognized faster than when shown to the right of fixation, and read from left to right; the reverse was found for normally oriented digits.

A familiar symbol on a printed page appears in essentially two dimensions of visual space, whereas the familiar contexts in which such symbols normally appear, such as books, are three dimensional objects. It is possible, therefore, either that the process of recognizing familiar two dimensional patterns is different from the process of recognizing familiar objects, that the spatial orientation of printed symbols is irrelevant to their recognition, or else in some sense that pattern recognition occurs as a special case of space perception.

One of the possibilities is implausible. Since it is obvious that an *S* can recognize a familiar symbol regardless of its spatial orientation, there may be features of such symbols which are invariant under all spatial transformations. If this notion were true, one ought to be able to easily locate and identify orientation invariant features of patterns. This is not an easy thing to do. It is easy, however, to discriminate between different spatial transformations of a symbol, providing that the figure is not symmetrical. This implies it is not invariances of the figure but the invariances of spatial transformations which are learned and which make it possible to recognize a familiar symbol regardless of its spatial orientation.

The two remaining possibilities seem equally plausible at first glance, primarily because the transformation of a printed symbol around its axis of depth can occur as a result of transforms in either depth or in two dimensional space. Consider, for example, a plastic digit 1 placed on a page, and appearing to *O* in its normal orientation (N). The digit might also be turned over, from side to side, so as to now appear as a mirror image of a normally oriented "1." The operation of turning the digit over in this way is equivalent geometrically to rotating the piece 180° around its *y*, or vertical axis (Y). However, we might have chosen instead to turn the digit over from top to bottom, thereby producing an inverted mirror image of the 1. This second type of operation is equivalent geometrically to rotating the digit 180° around its *x*, or horizontal axis (X). Fig. 1 shows how the digit would appear to *O*

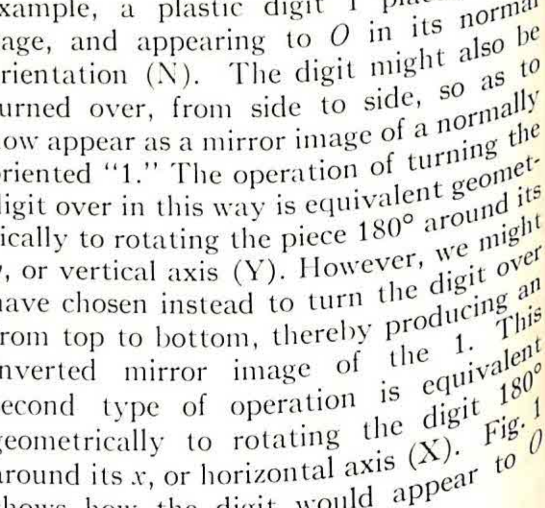


FIG. 1. The digit "1" as it appears normally, and after 180° rotation around its *x*, *y*, and *z* axis.

---

[1] This research was supported by Grants A-7811 and E-3329 from the National Research Council of Canada. The assistance of Calvin Peddle is gratefully acknowledged.

[2] Requests for reprints should be sent to R. L. Taylor, Department of Psychology, Memorial University of Newfoundland, St. John's, Newfoundland, Canada.

after either 180° *x* or *y* axis rotation. Both of these transformations are "legitimate," since they utilize the principal spatial dimensions of a printed page. The third possibility is "legal" only in three dimensional space, because it involves 180° rotation around the *z* axis, the axis of depth (Z). It is something one might do to a book. But a digit appearing after a Z transform is identical to a digit undergoing, in any order, *both* an X and Y transform. A simple proof of this assertion can be obtained by drawing an asymmetric figure on a piece of paper. In either sequence, turn the paper over top to bottom and again side to side. The result is the same as if the paper had simply been rotated 180° around its axis of depth. The fact that a person can recognize a printed symbol which appears after a Z transform, therefore, does not necessarily imply either that he knows the invariances of depth and is using that knowledge, or that he is processing the printed symbol in a page—relative two dimensional space.

When *S*s read lines of text from a page, they consistently make fewer errors and are faster with letters that have undergone a Z transform, but are slower and make errors most frequently with letters having undergone an X transform (Kolers, 1968; Kolers & Perkins, 1969a, 1969b). This result is surprising because if *S*s were decoding a Z transform in two dimensions by successively applying X and Y transforms, one would have expected that Z would be harder than either or both X and Y separately.

It is also possible that spatial orientation and the direction *S*s read may interact to determine speed of recognition. Kolers and Perkins (1969b) report that letters which "face" to the left are named faster when read from left to right, whereas letters that face to the right are named faster when read from right to left.

The purpose of this study was to further investigate the relative difficulty of spatially transformed symbols in a reading task. Instead of letters, however, only digits which present a unique pattern in each orientation are used as test stimuli.

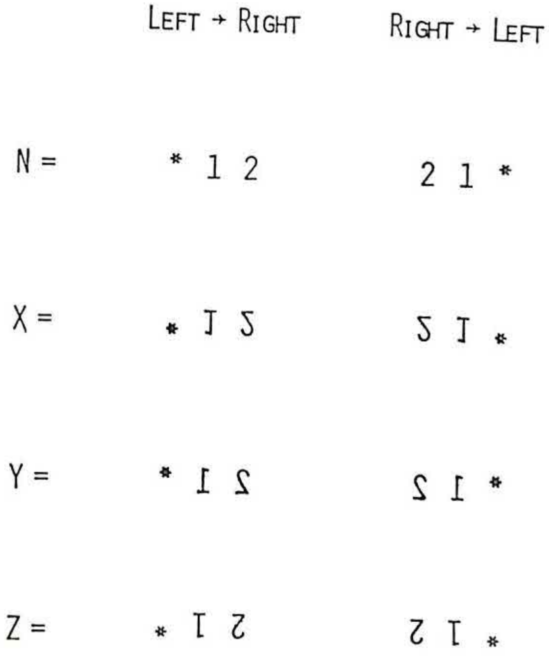


FIG. 2. The digit number, 12, as it would appear to an *S*. (The variables combined orthogonally are reading direction (left to right, right to left), signaled by the position of the asterisk, and spatial transform [Normal, or 180° rotation around *x*, *y*, or *z* axis].)

In addition, it seemed plausible that the direction figures are read might be a particularly important variable when the same visual patterns mean two different things, depending on how they are read. Thus, → 24 is twenty-four, but 24 ← is not.

METHOD

Design.—The experiment followed the basic plan shown in Fig. 2. There were two principal variables: spatial transform, represented by digits appearing either in their normal orientation (N), or after 180° rotation around their *x*, *y*, or *z* axis; and reading direction, represented by an asterisk appearing either to the left or right of the digits (an asterisk appearing on the left indicates that the digits should be read from left to right). There were 11 *S*s and each *S* was tested in all eight conditions. The order of conditions was determined randomly, and a mixed-list mode of presentation was used.

Apparatus.—Apparatus consisted of a tachistoscope equipped with automatic paper-tape stimulus changers (Scientific Prototype, 2-channel, with BCI changers), a general purpose digital computer (NOVA), and a millisecond timer (Hunter). The tachistoscope presented stimuli to *S*, who responded on each trial by pressing one of two levers. When a correct response was made, *E* entered *S*'s reaction

TABLE 1

Mean Digit Number Recognition Time (msec.) by Subject, Reading Direction, and Spatial Transform

S	Left to right				Right to left			
	N	X	Y	Z	N	X	Y	Z
1	563	657	632	638	604	625	610	586
2	838	977	925	926	894	901	871	862
3	682	746	728	751	690	731	709	684
4	699	792	778	762	707	720	699	672
5	502	564	548	528	520	527	520	507
6	660	757	755	806	711	725	722	670
7	607	664	635	635	602	609	592	609
8	603	695	697	671	653	673	691	632
9	642	726	721	720	683	716	726	661
10	916	1015	1013	934	976	991	981	892
11	773	943	928	891	902	914	913	856
$\bar{X}$	680	776	760	751	722	737	730	694

Note.—Abbreviations: N = normal orientation; X = rotation around the *x* axis; Y = rotation around the *y* axis; Z = rotation through the *z* axis.

time into the computer. Stimulus changers then incremented position and automatically started another trial.

Stimuli.—The stimulus sequence in one trial consisted of a word number (e.g., twenty-three) followed by a digit number (e.g., 17). Word numbers were shown in Channel 1, digit numbers in Channel 2 of the tachistoscope. Both types of stimuli were produced by the computer via teleprinter directly on paper tapes. The program which generated the stimuli used a random number function to produce a mixed sequence, subject to the following constraints: the digits 1, 2, 3, 4, 5, 7 in all possible pairs excluding twins formed the population of potential numbers; no number was permitted to follow itself within two trials; an equal number of trials occurred in which the word number and digit number agreed and disagreed; and the frequency of each combination of reading direction and type of spatial transformation for digits fell within 10% of all other combinations in the series.

Since Channel 1 of the tachistoscope produces a Y transform of the stimuli, relative to *S*, it was necessary to produce word-number tapes by using a carbon. Also, since a mixed-list design was used, it was necessary to run each digit-number tape through the teleprinter four times, varying feed direction and placement of a carbon. These tapes were then punch coded and placed in the automatic stimulus changers of the tachistoscope. Eight different sets of tapes were produced, representing eight different randomized sequences.

Subjects.—The *S*s were seven male and four female undergraduate students at Memorial University. All *S*s stated a right hand use preference, and each was paid for his time and services. The *S*s were volunteers.

Procedure.—All *S*s were tested individually. They were given instructions to respond as rapidly and as accurately as possible. On any given trial, *S* was shown a word number followed by a digit number. The *S*s' task was to press one lever if the digits described the same number as the words (YES) and to press another lever if the digits described a different number than the words (NO). Levers were assigned to hands, YES to *S*'s right hand.

An asterisk (*) appeared in the center of *S*'s field of view in the tachistoscope (see Fig. 2). The digits were presented either 2° to the left or right of the asterisk, with position signaling the way that the number was read correctly. The intertrial interval was fixed at 5 sec. Word numbers were exposed for 4 sec., while a response by *S* terminated the exposure of a digit number.

RESULTS

The overall error rate was 6.2%. Error trials were omitted from the analysis of the latencies. In order to obtain scores, the reaction time for each correct trial was stored in the computer as its natural logarithm and sorted by experimental condition. At the end of a session means were obtained on converted latencies by condition, reconverted by exponent, and printed out by the computer with analyses of *S*'s performance. A mean for each *S* in each condition was then obtained across the eight sessions. Since there was no indication that mean reaction time by hand interacted with any of the other conditions, hand data were pooled.

The *S*s consistently produced slower

reaction times for X transformed digit numbers, and generally were faster with Z transformed digit numbers. The same rank-order difficulty has been reported previously when *Ss'* task was to read letters from lines of text (Kolers, 1968; Kolers & Perkins, 1969a, 1969b). Also, the failure to find consistent differences between Y and Z in this task parallels reaction time data obtained by Shepard (1971), who had *Ss* decide whether two perspective line drawings portrayed objects of the same three-dimensional shape or not. It does not appear to be true in general, however, that familiar symbols are recognized easier when they face the same direction they are read: X transformed digit numbers, which "face" to the left, produced faster reaction times when read from right to left than when read from left to right by all *Ss* tested.

The *Ss*, with but 2 exceptions out of 44 possibilities, were faster with normally oriented digit numbers read from left to right and spatially transformed digits read from right to left.

DISCUSSION

It is possible to account for the principal findings in the following way. While reading normal text, one is most likely to have at one time or another tried to read symbols from a book that was upside down (Z); occasionally, one might have tried to read them from a mirror (Y); and very rarely would one have tried to read *through* a page that was upside down (X). This notion can be formalized in part by assuming four generating operators (N, X, Y, Z), where N is the identity operator, and a set of geometric operators (*). The task of decoding each spatial transformation is a two-stage process, consisting first of normalizing each symbol and then deciding on the identity of the resulting shape. Once the type of transformation underlying the appearance of a symbol has been identified, the symbol can be made normal (n) by reapplying the same 180° transform: thus, $N^*N \rightarrow n$; $X^*X \rightarrow n$; $Y^*Y \rightarrow n$; $Z^*Z \rightarrow n$. Differences in speed of recognition occur in response to differences in how long it takes to identify which transform a symbol has undergone, with differential familiarity being the critical variable. This pattern recognition device would work equally well with templates or topographic features lists. The principal difficulty is its implausibility: It assumes that *Ss* process symbols appearing on a flat surface as if they were three-dimensional objects.

A more parsimonious model assumes three primitive generating operators (N, Y, Z), implying that a decision about any three-dimensional transform can be made by normalizing across either one or two primitive operators. In order to decode an X transform, for example, $X^*Y^*Z \rightarrow n$. Since two primitive operations occur serially, the process takes longer than either primitive alone.

If *Ss* used the asterisk appearing with each digit number as a fixation point, as instructed, then all spatially transformed digit numbers were processed faster when they were projected to the right cerebral hemisphere, while normally oriented digits were processed faster when they were projected to the left cerebral hemisphere. It is tempting to speculate further, therefore, that the process of normalizing symbols for orientation is dependent on mechanisms in or driven by the right half of the brain.

REFERENCES

KOLERS, P. A. The recognition of geometrically transformed text. *Perception & Psychophysics*, 1968, **3**, 57–64.

KOLERS, P. A., & PERKINS, D. N. Orientation of letters and errors in their recognition. *Perception & Psychophysics*, 1969, **5**, 265–269. (a)

KOLERS, P. A., & PERKINS, D. N. Orientation of letters and their speed of recognition. *Perception & Psychophysics*, 1969, **5**, 275–280. (b)

SHEPARD, R. N., & METZLER, J. Mental rotation of three-dimensional objects. *Science*, 1971, **171**, 701–703.

(Received May 5, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 400–406

INTERFERENCE IN SHORT-TERM RETENTION OF DISCRETE MOVEMENTS[1]

A. S. FAUST-ADAMS[2]

University of Michigan

The retention of discrete lever movements was examined, with additional movements of the same lever interpolated in a 12-sec. retention interval. In Exp. I, the position within the retention interval of one or two interpolated movements was varied. The absolute error of recall was significantly greater as the amount of interpolated material increased, and as its position moved toward the end of the retention interval. In Exp. II, four types of movements, ranging from a complete active movement to a complete passive movement, were interpolated. Forgetting was directly related to the amount of motor output during the retention interval, indicating that the most important information used to encode a discrete movement is the motor output required to execute it.

Williams, Beaver, Spence, and Rundell (1969) have shown that retroactive interference effects can be demonstrated in the short-term retention of discrete motor movements. Such movements, however, may be seen as involving a number of distinct components, such as the motor output required for the movement or the feedback the movement generates. The present experiments were designed to determine how much of the total interference effect is attributable to each of these components. It can then be argued that the original movement is encoded in terms of those components which produce the greatest interference effects.

EXPERIMENT I

The original design for an experiment to compare various interference effects involved interpolating different movement components at different positions within the retention interval. It was first necessary, therefore, to determine whether position of interpolation itself had any differential effect on retention in this situation.

Method

Apparatus.—The apparatus used was the same in both experiments. Fitted to a right-handed student's desk chair was a lever which could rotate in a horizontal plane through an angle of 130°. In its left-most position, the lever was parallel to *S*'s frontal plane. It was equipped with an elbow support and an adjustable vertical bar, which *S* gripped with his hand. In Exp. I, the lever was always moved by *S*. In Exp. II, it was sometimes moved by means of a bidirectional motor and a clutch. In either case a brake stopped the lever's movement when necessary. With the clutch engaged and the motor on, the arm rotated at 54 rpm, and under these circumstances a torque of 15 kg.-cm. was sufficient to cause the clutch to slip and the arm to stop moving. There was a shield over the lever so that *S* was not able to see either the lever or his arm.

Instructions were presented to *S* on a cathode-ray display controlled by a PDP-1 computer. The computer was also used to control the motor power and direction, the clutch and the brake, and to monitor the position of the lever.

Design.—Each *S* was presented with 96 trials, comprising eight experimental conditions, six target angles, and two replications, in a factorial design. The trials were presented in random order, with the exception that one replication was completed before the other was begun. Any difference between the replications therefore constitutes a practice effect. The target angles which were used ranged from 15° to 40° in 5° steps.

A preliminary experiment revealed that the size of the target angle and the direction of its presentation had significant effects on algebraic error, but none of the interactions involving these factors were of interest. In the present experiment, the Angles × Direction of Presentation and Angles × Direction of Recall interactions were therefore confounded with the *S*s effect. Within each *S*, each angle was paired, in a balanced incomplete-block design, with only one of the four possible combinations of the two directions of presentation and the two directions of recall. Across every four *S*s, however, all possible combinations occurred equally often.

---

[1] This report is based on a dissertation submitted by the author to the Horace H. Rackham School of Graduate Studies, University of Michigan, in partial fulfillment of the requirements for the PhD degree. The research was supported by the Advanced Research Projects Agency, Department of Defense, and monitored by the Air Force Office of Scientific Research, under Contract AF 49(638)-1736 with the Human Performance Center, Department of Psychology, University of Michigan. Reproduction in whole or in part is permitted for any purpose of the United States Government. The author is indebted to A. W. Melton for his advice and encouragement, and for his critical reading of a draft of this paper.

[2] Requests for reprints should be sent to A. S. Faust-Adams, who is now at the School of Psychology, University of New South Wales, Kensington, New South Wales 2033, Australia.

The retention interval in this experiment was 12 sec., permitting four interpolated movements, one in each of four 3-sec. intervals into which the retention interval was conceptually divided. In each of the eight different treatment conditions, the retention interval was filled differently. In Cond. 1, the retention interval was unfilled. In Cond. 2, 3, and 4, it was filled with just one interpolated movement. In Cond. 2, this movement occurred in the first position, i.e., in the first of the four 3-sec. intervals. In Cond. 3 it occurred equally often in each of the two middle positions, and in Cond. 4 it occurred in the last position. In Cond. 5, 6, and 7, two movements occurred in the first two, the middle two, and the last two positions, respectively. Finally, in Cond. 8, the retention interval was completely filled with four interpolated movements. The various ways in which the retention interval was filled are shown schematically in Table 1.

The construction of each trial.—For each trial, the starting point for the target movement was chosen from within a 90° "working range," placed in the center of the lever's 130° movement range. There was, therefore, an additional 20° of movement at either end before the lever struck against the fixed stops. The starting point for all movements was chosen randomly, but with the restriction that if each movement was made accurately the lever would never move outside the working range.

In Cond. 2 through 8, interfering movements were interpolated between presentation and recall, each movement taking approximately 3 sec. The finishing point for each of these movements was chosen randomly from the working range, with the restriction that each movement should be at least 5° in extent. The starting position for recall was always chosen with the restriction that *S* could not predict the direction of recall.

Procedure.—Each trial, including the first, began with the display of the word REST for 8 sec. During this period the lever was set to the starting position for that trial. At the end of this period, the words GRASP LEVER were displayed for 4 sec., and *S* was instructed to rest his arm on the lever at that point. The words MOVE AND REMEMBER then appeared, with an arrow beneath them indicating to *S* the appropriate direction for him to move. On this signal, *S* moved the lever in the direction of the arrow until the brake came on.

For Cond. 1, and for the rest portions of the retention interval in Cond. 2 through 7, the screen remained blank, and *S* was instructed to keep his arms still until some instruction appeared. When an interpolated movement was to be presented, a number and an arrow appeared on the screen at the beginning of the 3-sec. interval. The number represented the distance in degrees through which *S* was to try and move the lever, while the arrow indicated the direction. After this display had been on for 1 sec., the word MOVE appeared above it. The *S* was instructed not to move until this appeared. Once *S* started moving, the lever's position was read every 200 msec., and the end of the movement was defined as having occurred if the lever's position did not change by more than 2° in any 200-msec. period. The *S* was allowed 2 sec. to complete the movement, making 3 sec. in all for each interpolated event. Instructions stressed that each movement had to be made smoothly and deliberately and that corrections could not be made once *S* stopped moving.

TABLE 1

LOCATION OF THE INTERPOLATED MOVEMENTS WITHIN THE RETENTION INTERVAL FOR EACH OF THE TREATMENT CONDITIONS, WITH THE ASSOCIATED ALGEBRAIC ERROR AND ITS VARIANCE, EXPERIMENT I

Condition	3-sec. interval[a]				Algebraic error	Variance
	1st	2nd	3rd	4th		
1					−.92	58.98
2	X				−1.24	61.64
3		X - - -	X[b]		−.73	61.12
4				X	.95	75.60
5	X	X			−.16	57.16
6		X	X		−.39	65.01
7			X	X	.12	84.68
8	X	X	X	X	−.16	80.04

Note.—Algebraic error is measured in degrees.
[a] An X indicates the presence of interpolated material.
[b] Material is divided between these two positions.

At the end of the retention interval the word RECALL was displayed, together with an arrow indicating the direction. This was a signal for *S* to try to move the lever through the target angle, 3 sec. being allowed for him to complete the movement. Finally, the word REST appeared, which was a signal for *S* to take his arm off the lever and rest it in his lap.

Whenever *S* made a mistake by failing to follow directions, a signal such as WRONG DIRECTION appeared for 2 sec. to indicate the nature of the mistake. This was followed by REST, then by the beginning of the next trial. The aborted trial was then automatically repeated at the end of an arbitrary block of trials in which it occurred.

Each *S* was given two short practice periods. The first period continued until *S* had completed five consecutive trials without any mistakes, such as moving in the wrong direction. Instructions were then given on the bonus system, which was 3¢ each time *S* reproduced the angle to within ±2° of its correct value. During the second practice period, *S* was told after each trial whether or not he had attained the criterion, and the practice ended when he

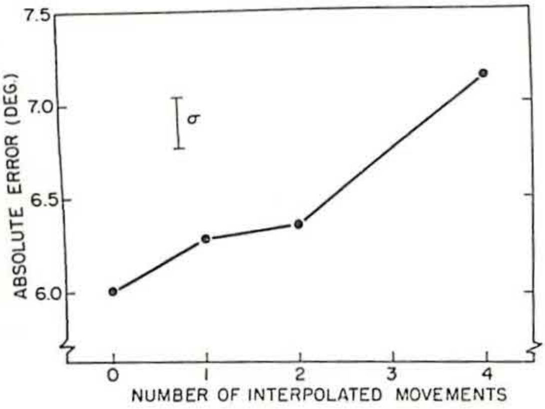


FIG. 1. Absolute error as a function of amount of interpolated material.

had attained it three times, or had completed 10 trials, whichever took longer.

Subjects.—The *S*s were 28 right-handed males who had volunteered to serve in paid experiments. They were paid at the rate of $1.50 per hour, plus a bonus based on performance. All were naive to motor-retention experiments.

Results and Discussion

The dependent variables were the absolute and algebraic errors to the nearest degree. The experiment involved a total of 2,688 trials, of which 75, or 2.8%, were missing. Although trials on which *S* made an error were repeated, there were some occasions when the paper tape was misread, with the result that an impossible trial was attempted. In such cases, the program automatically continued with the next trial. Since there was no case in which a trial was missing from both first and second replications, each missing trial was simply replaced with the corresponding value from the other replication, and the degrees of freedom were reduced where appropriate.

A four-way analysis of variance was carried out on the absolute errors, the factors being *S*s, practice, treatment conditions, and angles. The treatment conditions will be discussed below. Although the practice effect was not significant, the angle effect was significant, $F\ (5,\ 135) = 7.30$, $p < .001$. There were significant interactions of *S*s with treatments, $F\ (189,\ 870) = 1.39$, and of *S*s with angles, $F\ (135,\ 870) = 3.67$, both ps $< .01$. These reflect greater variability across treatments and angles for some *S*s than others.

The overall effect of amount of interpolated activity is shown in Fig. 1. The data have been averaged across Cond. 2, 3, and 4 to give the mean effect for one interpolated movement, and across Cond. 5, 6, and 7 for two movements. The significance of the effect of number of interpolated movements was tested with a planned comparison in which weights of -3, -1, -1, -1, 1, 1, 1, and 3 were used for the eight treatment means in order. The results were significant at the 5% level, $F\ (1,\ 189) = 6.54$, showing a direct increase in absolute error with an increase in the number of interpolated movements, the retention interval remaining fixed.

The effect of varying the position of interpolation within the retention interval is shown in Fig. 2. In order to test this effect, separate analyses of variance were carried out for one and for two interpolated movements. There was a significant effect of the position of the two interpolated movements, $F\ (2,\ 54) = 5.51$, $p < .01$, but not for one movement, $F\ (2,\ 54) = 1.32$. No interactions reached significance in either of these analyses.

Since Bilodeau (1966) has suggested that the variance of errors will increase as a result of forgetting, the variances associated with the treatment conditions were examined and are given in Table 1. It is not appropriate to carry out a series of pairwise F tests on these results, but Cochran's C test for homogeneity of variance (Winer, 1962, p. 94) was carried out. For the three variances involving one interpolated item (Cond. 2, 3, and 4), $C = .38$, and for two items (Cond. 5, 6 and 7), $C = .41$. Both of these are significant at $p < .05$, indicating significantly nonhomogeneous variances. An examination of the individual variances indicates that, with the exception of Cond. 2 and 3, which are almost identical, there is an increase in the variance of the recall as the interpolated material moves toward the end of the retention interval.

No immediate explanation of this is apparent, but one possibility is that the closer the interpolated material is to the recall, the greater the likelihood of response competition between the interpolated and retained movements. If this is so, there should be some relationship between the similarity of the interpolated and retained movements and recall performance. Alternatively, there might be some relation between the angles recalled and the interfering items, which would be analogous to finding intrusions in the verbal situation.

These two possibilities were examined for trials in Cond. 7, where two interpolated movements appeared at the end of the retention interval. The first possibility was examined by calculating the regression of the absolute difference between the angle presented and each of the two interpolated angles. In all cases the angle actually executed was used. The multiple-regression coefficient was .09, which is not significant. The second possibility was examined by calculating the regression of the angle recalled on the two interpolated angles. The result was not significant, the coefficient being .08. Finally, the regression of absolute error on the sizes of the two interpolated angles was calculated, with the aim of examining the hypothesis that larger interpolated angles might produce larger errors, without regard to the sign of the error. The regression, with a coefficient of .07, was again not significant.

EXPERIMENT II

The results of Exp. I suggest that the size of the interpolated angles does not affect retention in this situation. Another possibility is that the nature of the interpolated movement itself has some effect. It is known that an active movement of the same general type as that being retained will cause an increase in forgetting, while at the other extreme, neither the writing nor the mental activity involved in an interpolated paper-and-pencil task has any effect (Williams et al., 1969). The present experiment examined this dimension by interpolating movements which involved various components of the retained movement.

Four components were chosen as being both tractable and theoretically interesting.

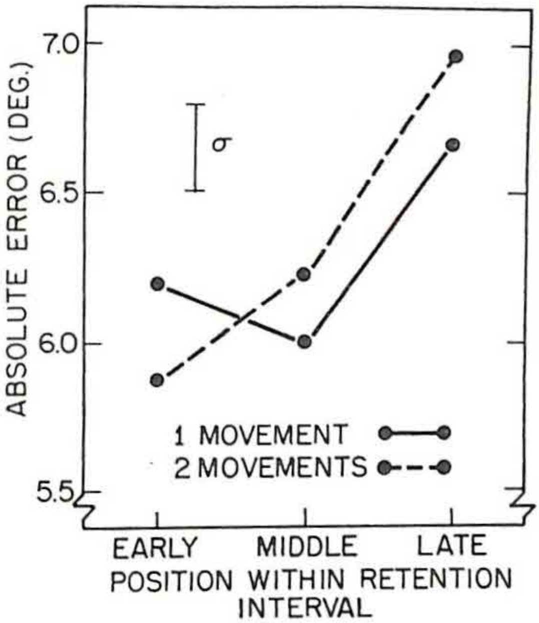


FIG. 2. Absolute error for one and two interpolated movements as a function of the position of the movements within the retention interval.

The first is the preparation or planning of a movement. Welford (1968, p. 140) suggests that brief movements such as the present discrete motor movement are "ballistic," in that they are initiated and carried out as a single unit, with *S* monitoring feedback only from the beginning and end of the movement.

This preparation component was manipulated by interpolating some items in which *S* was given the angle and direction to move through, but then told DON'T MOVE instead of MOVE. This will be referred to as a Movement Preparation item. In the previous experiment, the angle and direction were displayed for 1 sec. prior to the MOVE signal. The same applied in this case, and it was again stressed that *S* should move as quickly as possible if MOVE appeared. It is therefore assumed that when DON'T MOVE appeared, *S* used the prior 1-sec. period to prepare the movement, even though it was not actually executed.

The second component was the actual initiation of the movement, including whatever feedback might result from such initiation. The item which involved this component, the Movement Initiation item, also involved the first component mentioned above. In other words, *S* was asked to prepare and initiate the movement, but once he moved 2° from the starting point the apparatus took over and finished moving his arm through the displayed angle.

In performing both of these interpolated items, *S* was to be exposed for at least 1 sec. to a display which indicated in degrees the extent of the movement. If both of these two items are just as detrimental to recall as the complete movement used in the previous experiment, the display of this number may be responsible. The *S* may transform the number into his own measure of the extent of the movement, and this in itself may be sufficient to affect the retention of the original item. The possibility that such an effect may be important was checked by including an item identical to the second, except for the absence of the number indicating the movement extent. This will be referred to as a Movement Initiation without Display item. For this item, *S* was required to initiate the movement, but he had no idea how far the machine would move his arm once it took over.

The final component which was examined involved all those aspects of the movement that are not involved in the initiation and execution by *S*. This component will be termed the feedback component, but it would be best to define it operationally, in terms of the Passive Movement item which was used to measure it. The *S* saw the angle and direction display, but at the end of the 1-sec. period the word RELAX appeared on the screen, and the apparatus moved *S*'s arm through the displayed angle. The *S* was therefore subjected to much of the feedback involved in making the complete movement.

Method

Design.—Details of the design which are not mentioned were the same as in Exp. I. Each *S* received 84 trials made up of 7 Conditions × 6 Angles × 2 Replications. The 7 conditions differed only in the ways in which the retention intervals were filled. They are as follows: In Cond. 1 there were four Movement Preparation items; in Cond. 2, four Movement Initiation items; in Cond. 3, four Movement Initiation without Display items; and in Cond. 4, four Passive Movement items. In Cond. 5 there were four Active Movement items, exactly as in Cond. 8 of Exp. I. Cond. 6 was a rest condition in which the retention interval was unfilled, exactly as in Cond. 1 of Exp. I. Finally, Cond. 7 was a control condition in which Active Movement and Movement Preparation items were mixed in equal numbers. Without Cond. 7, *S* would have known that one Movement Preparation item always indicated that three more were to come, and he would not have prepared the subsequent movements.

Procedure.—Changes in the procedure from Exp. I were relevant to the new items used. If *S* resisted during any of the conditions involving passive movement, so that the clutch slipped and the lever did not reach the finishing point by the time the 3 sec. were up, the error signal YOU MUST RELAX was displayed. A Movement Initiation item was signaled by the appearance of a diamond around the number indicating the size of the angle, while for a Movement Initiation without Display item, the diamond appeared, but there was no number inside it.

Three practice periods were given, introducing *S* gradually to the various conditions.

Subjects.—There were 32 right-handed male *S*s drawn from the same source as for Exp. I. All were again naive to motor-retention experiments, and they were given the same base pay and bonus as before.

Results and Discussion

Of the 2,688 trials, a total of 101, or 3.6%, were missing because of apparatus failure. These were estimated, as in Exp. I, by replacing each with the corresponding value from the other replication, and the degrees of freedom were reduced where appropriate.

In an analysis of variance on the absolute errors, two of the main effects were significant: angles, F (5, 155) = 10.74, $p <$ .001, and treatment conditions, F (6, 186) = 2.34, $p <$ .05. The treatment conditions effect was subjected to an F test, since no contrasts were planned on the means. There were two significant interactions: *S*s × Angles, F (155, 829) = 2.06, $p <$ .001, and Treatment × Angles, F (30, 829) = 2.78, $p <$ .001, for which the explanations are as previously given.

An examination of the algebraic errors in both the experiments indicated that the same main effects were significant in all the absolute-error and algebraic-error analyses. There were many significant interactions in the algebraic-error analyses, making interpretation of the main effects difficult. The angle effects, however, were highly consistent in both experiments (in each case $p < .001$), and are shown in Table 2. In each case there is undershooting for the three largest angles and overshooting for the three smallest.

Table 3 shows the treatment condition means for the absolute error. Since the treatment condition means were not on any type of scale, and there were no specific prior hypotheses about their order, posterior comparisons were carried out using Duncan's new multiple-range test (Duncan, 1955), the results of which can be summarized schematically as follows: $\underline{4162\underline{37}}\underline{\underline{5}}$. Treatment means not underlined by a common line differ significantly, the significance level being 5% or better.

It is clear that Cond. 5 (Active Movement) produced recall which is significantly worse than almost all the other conditions. The difference between the means for Cond. 3 (Movement Initiation without Display) and Cond. 5 was .737°, while the difference required for significance at the 5% level was .745°. This suggests that Cond. 2 and 3 (the two Movement Initiation conditions), while not significantly different from each other, have together a significantly smaller effect than Cond. 5 (Active Movement). Removing from the display of Cond. 2 (Movement Initiation) the number which tells *S* how far he will be moving clearly has no effect on the results. Similarly, simply asking *S* to prepare a movement without actually making one has no significantly detrimental effect on recall. In fact, both Cond. 1 (Movement Preparation) and Cond. 4 (Passive Movement) produced mean errors slightly below that of Cond. 6 (Rest).

The interpolation of passive movements, as shown by the results of Cond. 4, obviously has no detrimental effect on recall. The kinesthesis generated by these passive movements would be very similar to the kinesthetic feedback produced when *S* makes an active movement. The fact that no significant interference was produced by the passive movement therefore suggests that under normal circumstances *S* does not remember the angle he moves through in terms of any transformation of the feedback it generates. This is in keeping with physiological evidence indicating that a response can be learned without kinesthetic feedback, and that a consistent response can be executed without such feedback (Lashley, 1917; Laszlo, 1967; Taub, Bacon, & Berman, 1965).

TABLE 2

MEAN ALGEBRAIC ERRORS FOR THE TARGET ANGLES, EXPERIMENTS I AND II

Exp.	Angle					
	15°	20°	25°	30°	35°	40°
I	2.35	2.43	.39	− .81	−2.47	−3.78
II	2.50	2.31	.29	−2.32	−3.31	−3.95

Note.—Algebraic errors are measured in degrees.

TABLE 3

MEAN ABSOLUTE AND ALGEBRAIC ERRORS AND ALGEBRAIC ERROR VARIANCES FOR THE TREATMENT CONDITIONS, EXPERIMENT II

Treatment condition	Absolute error	Algebraic error	Variance
1. Movement preparation	5.68	− .58	52.90
2. Movement initiation	5.86	−1.01	55.87
3. Movement initiation without display	5.91	− .49	57.79
4. Passive movement	5.60	.57	51.98
5. Active movement	6.65	− .15	77.27
6. Rest	5.76	−2.60	47.35
7. Control	6.17	− .97	61.44

Note.—Algebraic errors are measured in degrees.

The variances for the algebraic errors shown in Table 3 were examined with Cochran's test, giving $C = .19$, $p < .01$. The variances are in the same order as the mean absolute errors, except for Cond. 6 (Rest), which has the smallest variance.

It seems that the important variable affecting retention is the amount of active movement made by *S* during the retention interval, since only interpolated items involving complete active movements have a significantly detrimental effect on recall. Further support for this conclusion is given by the fact that Cond. 2 and 3 (Movement Preparation) and Cond. 7 (Control) appear to produce forgetting which is intermediate between that produced by Cond. 6 (Rest) and Cond. 5 (Active Movement). In both of these conditions there is active movement, but in each case it is less than in Cond. 5.

Pepper and Herman (1970) have examined the question of what happens to the trace of a motor act as it is forgotten, directing their attention to the effects on algebraic errors. In a number of studies on the retention of a movement extent (Adams & Dijkstra, 1966; Posner, 1967; Stelmach, 1969), Pepper and Herman find a negative time error, i.e., an undershooting, which increases with time, a finding which they account for in terms of a fading trace theory. A second finding with respect to the algebraic effect is that any interpolated material tends to result in a more positive error. The results of the present experiments support this finding. In Exp. I, a slight undershooting of about 1° is reduced to zero with interpolated material ($.05 < p < .01$, see Table 1). In Exp. II, a mean algebraic error of about $-2.5°$ is raised to between $-1°$ and $.5°$ by the interpolation of any material ($p < .05$, see Table 3). Although the large between-*S* differences and the many interactions make it quite clear that these are only average effects, they appear to be quite consistent. Pepper and Herman offer an explanation in terms of the distracting effects of the interpolated material.

An explanation must also be given, however, for the increase in absolute error which is found in this and other studies. The fact that the increase in absolute error is accompanied by a variance increase suggests that *S* is trying to match a trace which is not only "shrinking" (a less ambiguous term than "fading"), but also getting dimmer in some sense, so that it is reproduced less accurately. This suggestion is strongly supported by the effect of angles on the algebraic error, as shown in Table 2. The finding that the smaller amplitudes in a series are overestimated and the larger ones underestimated is not new; is was explored at length by Hollingworth (1909). It is likely that this represents a tendency for *S*s to err toward the mean whenever they are in doubt as to the size of the angle presented, an explanation which avoids reference to a spatially extended image. This effect is obviously much stronger than any absolute "shrinkage." If it were not, we would expect the negative time error to be present for each angle taken separately, and not just for the mean of all angles.

REFERENCES

Adams, J. A., & Dijkstra, S. Short-term memory for motor responses. *Journal of Experimental Psychology*, 1966, **71**, 314–318.

Bilodeau, E. A. Retention. In E. A. Bilodeau (Ed.), *Acquisition of skill*. New York: Academic Press, 1966.

Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–14.

Hollingworth, H. L. The inaccuracy of movement. *Archives of Psychology*, 1909, **2**, 1–87.

Lashley, K. S. The accuracy of movement in the absence of excitation from the moving organ. *American Journal of Physiology*, 1917, **43**, 169–194.

Laszlo, J. I. Training of fast tapping with reduction of kinaesthetic, tactile, visual and auditory sensations. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 344–349.

Pepper, R. L., & Herman, L. M. Decay and interference effects in the short-term retention of a discrete motor act. *Journal of Experimental Psychology*, 1970, **83**(2, Pt. 2).

Posner, M. I. Short-term memory systems in human information processing. *Acta Psychologica*, 1967, **27**, 267–284.

Stelmach, G. F. Prior positioning responses as a factor in short-term retention of a simple motor task. *Journal of Experimental Psychology*, 1969, **81**, 523–526.

Taub, E., Bacon, R. C., & Berman, A. J. Acquisition of a trace-conditioned avoidance response after deafferentation of the responding limb. *Journal of Comparative and Physiological Psychology*, 1965, **59**, 275–279.

Welford, A. T. *Fundamentals of skill*. London: Methuen, 1968.

Williams, H. L., Beaver, W. S., Spence, M. T., & Rundell, O. H. Digital and kinaesthetic memory with interpolated information processing. *Journal of Experimental Psychology*, 1969, **80**, 530–536.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received May 8, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 407–415

INTERSTIMULUS INTERVAL AND TIME ESTIMATION IN RATINGS OF SIGNALED SHOCK AVERSIVENESS[1]

MILTON D. SUBOSKI,[2] TONNAR G. BRACE,
Queen's University at Kingston,
Kingston, Ontario, Canada

LOUISE A. JARROLD,
Sir George Williams University,
Montreal, Quebec, Canada

KURT J. TELLER, AND RICHARD DIETER
Virginia Polytechnic Institute and State University

The hypothesis that anticipatory conditioned responses (CRs) can modify the aversiveness of an electric shock UCS was tested by having *S*s rate the aversiveness of signaled and unsignaled shock. Ratings of shock intensity were a U-shaped function of interstimulus interval (ISI) similar to the inverted U-shaped function for CRs in classical conditioning. Requiring *S*s to estimate the duration of the ISI yielded ratings of signaled shock below unsignaled at all ISIs, presumably because of more accurate placement of anticipatory CRs which attenuate the aversiveness of the shock. At a short ISI, differential ratings of signaled and unsignaled shock developed gradually over trials, as do CRs in classical conditioning. Ratings of signaled and unsignaled shock at long ISIs were similar to ratings at short ISIs when the ISI was gradually increased across trials. Thus, ratings of shock aversiveness can be "shaped" just as can high levels of CRs in classical conditioning and high levels of accuracy in time estimation.

According to a law-of-effect hypothesis (e.g., Perkins, 1968; Prokasy, 1965), the occurrence of conditioned responses (CRs) in classical conditioning is dependent upon the instrumental effects of the CR in modifying the sensory consequences of the unconditioned stimulus (UCS). A difficult problem for this interpretation is the fact that excellent conditioning of a wide variety of responses can be obtained with the use of electric shock as the UCS (e.g., Gormezano & Fernald, 1971). No mechanism whereby anticipatory CRs can attenuate the noxiousness of an electric shock UCS is readily apparent. Nevertheless, the hypothesis that CRs have instrumental consequences in modifying the effects of an electric shock UCS yields a number of predictions of interest to the present research. First, signaled shock should be preferred to, and rated as subjectively less intense than, unsignaled shock. Second, parameters that determine classical conditioning levels should have comparable effects on preference for signaled shock and subjective ratings of shock intensity.

A number of studies show that both rats (e.g., Perkins, Seymann, Levis, & Spencer, 1966) and humans (e.g., Badia, Suter, & Lewis, 1967) consistently select signaled over unsignaled shock in a choice situation. In contrast to the pronounced preference shown by *S*s for signaled shock, corresponding ratings of low aversiveness for signaled shock have not been consistently obtained. Signaled shock has been reported to be less aversive than unsignaled shock (e.g., Kimmel, 1967; Pervin, 1963), but more often signaled shock has been found to be more aversive than unsignaled (e.g., Furedy & Doob, 1971a, 1971b; Kimmel, 1967). The present research investigated the possibility that failures to obtain ratings of

---

[1] This research was supported by Grant A0-244 from the National Research Council of Canada and funds from the Department of Psychology of Virginia Polytechnic Institute and State University. Experiment I was presented at the meeting of the Eastern Psychological Association, Boston, April 1972. Experiment II was adapted from a thesis submitted by the third author in partial fulfillment of the requirements for the master's degree at Sir George Williams University. Experiment IV was presented at the meeting of the American Psychological Association, Honolulu, September 1972. A brief report of this research was presented at the meeting of the International Congress of Psychology, Tokyo, August 1972.

[2] Requests for reprints should be sent to Milton D. Suboski, Department of Psychology, Queen's University, Kingston, Ontario, Canada.

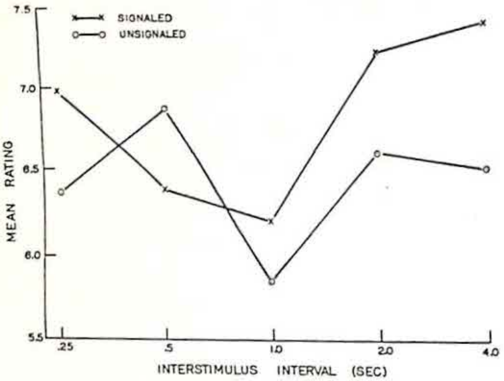


FIG. 1. Mean ratings of the intensity of signaled and unsignaled shock as a function of ISI for signaled shock (Exp. I).

signaled shock as less aversive than unsignaled were due to failures to use parameters favorable to classical conditioning.

The interstimulus interval (ISI) between conditioned stimulus (CS) and UCS is an important determinant of classical conditioning levels. Prokasy (1965) has suggested that the inverted **U**-shaped relationship between CR levels and ISI is a product of differential opportunity for CR–UCS overlap. According to this analysis, the poorer conditioning that results at postoptimal ISIs arises from the inability of Ss to accurately estimate longer time intervals and consequently time the CR to occur close to the UCS so that the reinforcing properties of CR–UCS overlap are maximized. This reasoning predicts that preference for signaled shock should be an inverted **U**-shaped function of ISI and intensity ratings should be a **U**-shaped function of ISI. The most informative data on this question are those provided by Kimmel (1967). She found that Ss judged shock at .5-sec. ISI as considerably less intense than did Ss receiving a 5-sec. ISI or unpaired conditioned stimuli (CSs) and UCSs. Furthermore, the 5-sec. ISI groups judged unsignaled shock on test trials as significantly *less* intense than the preceding signaled shocks, whereas the .5-sec. ISI group judged the unsignaled shock test trials as *more* intense than signaled shock. Thus Kimmel's data in particular suggest that ISI is an important determinant of the ratings of intensity of signaled shock. The present experiments represent a systematic investigation of the role of ISI in determining judgements of intensity or aversiveness of signaled shock.

EXPERIMENT I

In Exp. I, subjective judgments of shock were obtained from different groups of Ss as a function of ISI. If anticipatory CRs can modify the consequences of an electric shock, then the lowest ratings of subjective intensity of signaled shock should occur at the ISI most favorable to the development of CRs. For a variety of classically conditioned skeletal responses, the optimal ISI appears to be approximately .5 sec. (Gormezano & Moore, 1969).

Method

Subjects.—The *Ss* were 10 female and 40 male undergraduates recruited from psychology courses at Virginia Polytechnic Institute and State University.

Apparatus.—The apparatus consisted of a Model 504A Lafayette timer, a Model 50807 Lafayette stop clock, and a constant-current shock source. The *S* sat in a student chair and had a concentric disc electrode (Tursky, Watson, & O'Connell, 1965) mounted on the dorsal surface of the right arm, just proximal to the radial tuberosity. The *S* faced, at a distance of about 1 m., a flat-black plywood panel, 74 × 80 cm., with an unshielded General Electric NE 47 neon bulb centered in the panel, approximately at eye level.

Procedure.—Two female and eight male *Ss* were randomly assigned to each of five groups. The groups differed in the ISI used on signaled shock trials: .25, .5, 1.0, 2.0, or 4.0 sec. Each *S* received three preparatory shocks at intensities of .5, 1.0, and 1.5 ma. (by instructions to be rated 1, 5, and 9 on an arbitrary intensity scale) in an effort to provide anchor points, and then 48 more shocks at 1.0 ma. Each block of six trials contained one unsignaled and five signaled shocks. The unsignaled shock was randomly assigned to one of Trials 4, 5, or 6 of each block.

The ISI was adjusted using the stop clock, which was then disconnected. The *S* was admitted to the experimental room, the electrode was attached, and instructions were read which explained the use of the rating scale. The *Ss* were instructed to rate each shock by saying a number from one to nine after the shock and to watch the light during the experiment, although no mention was made of any association between the light and shock.

The *E* and the apparatus were behind the panel. Signaled trials were initiated by *E* with the hand switch on the timer, which made a soft onset click, audible to *S*. Unsignaled shocks were delivered by means of a silent mercury switch. The intertrial intervals were randomly selected from 10, 15, and 20 sec. The shock duration was .5 sec. and the offset of the neon bulb was coincident with shock onset on signaled trials.

Results

Figure 1 displays mean ratings of signaled and unsignaled shock as a function of ISI. Ratings of signaled shock were a **U**-shaped function of ISI, but the lowest ratings occurred at the 1.0-sec. rather than at the .5-sec. interval. Ratings of unsignaled shock did not appear to vary systematically with ISI, but were consistently below the ratings of signaled shock except at the .5-sec. interval. Analysis of variance of the mean ratings of signaled shock and the rating of the unsignaled shock for each block of six trials revealed that signaled shock was rated as significantly more aversive than unsignaled shock, $F\ (1,\ 45) = 4.06$, $p < .05$. Neither ISI nor the interaction between ISI and signaled versus unsignaled shock was significant. However, because of the considerably higher variance of the unsignaled shock ratings, a separate analysis of variance which included mean ratings of only signaled shock was conducted. This revealed a significant effect of ISI, $F\ (4,\ 45) = 2.66$, $p < .05$. Trend comparisons showed that only the quadratic component was significant, $F\ (1,\ 45) = 5.32$, $p < .05$.

Ratings of signaled shock intensity were the **U**-shaped function of ISI expected from the hypothesis that intensity ratings of signaled shock are determined by the same processes as are CR levels in classical conditioning. These data also agree with those of Kimmel (1967) on the effects of ISI on ratings of signaled and unsignaled shock. Kimmel found that signaled shock was judged as less intense than unsignaled at a .5-sec. ISI and more intense at a 5-sec. ISI. The same results were obtained in Exp. I for ISIs of .5 and 4 sec.

Overall, the outcome of this experiment was not very favorable to the hypothesis that classical conditioning with a shock UCS depends upon instrumental modification of the effects of the UCS by CRs. Signaled shock was rated as less intense than unsignaled only at the .5-sec. ISI, and even there the effect was not very large. The clearest effect, that signaled shock was rated as more intense than unsignaled shock, is directly opposed to the instrumental consequences hypothesis. This conclusion, however, fails to take into account any possible role of preparatory CRs. One possibility is that signaled shock is felt as more intense than unsignaled in the absence of any preparatory CRs (for as-yet-undiscovered reasons) and less intense when preparatory CRs occur. Consequently, the general design of Exp. I was repeated for Exp. II. A number of procedural improvements were made in order to make sure that the difference in ratings of signaled and unsignaled shock was not idiosyncratic to the procedures of Exp. I. The most important change, however, was the addition of a time-estimation task in which *S*s were instructed to press a key just before the shock. The key press provided one indication of trials on which *S* made an anticipatory response.

Experiment II

Method

Subjects.—The *S*s were 25 female and 25 male introductory psychology students from Queen's University, participating in the experiment as part of the course requirement.

Apparatus.—The *S* sat at a table facing a flat white panel at a distance of approximately .5 m. The top of a General Electric NE 40 bulb was exposed through a hole in the center of the panel at about eye level. The disc electrode was mounted on *S*'s left arm. A telegraph key was on the table in a position convenient to *S*'s right hand. Two Hunter timers, a Hunter Klockounter and the constant-current shock source were located in an adjacent room, connected by intercom to the *S* room.

Procedure.—Five male and five female *S*s were randomly assigned to each of five ISI groups: .25, .5, 1, 2, and 4 sec. The *S* was admitted to the experimental room, the electrode was attached, and *S* was given instructions to read. The instructions informed *S* that illumination of the light bulb would be followed by shock and required *S* to attempt to press the key just before each shock. The *S* was also given preliminary instructions on the use of a 7-point scale of unpleasantness and was told that the key press did not affect the shock.

From the control room *E* administered two series of unsignaled 100-msec. shocks of 2.5, 1.5, and .7 ma. The *S* was told that the 2.5-ma. shock was the most unpleasant that would be delivered and should be rated 7 on the scale. Similarly, *S* was in-

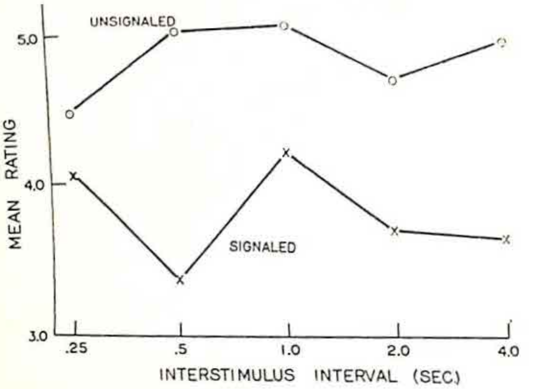


FIG. 2. Mean ratings of the unpleasantness of signaled and unsignaled shock as a function of ISI for signaled shock (Exp. II).

structed to rate the 1.5- and .7-ma. shocks as 4 and 1, respectively, with the .7-ma. shock represented as the least unpleasant that would be delivered. The *S* then received three presentations of the three shock intensities in a random order and *E* provided immediate feedback on the correctness of *S*'s ratings. Each *S* then received 25 signaled and 25 unsignaled 100-msec. shocks of 1.5 ma., ordered according to a Gellerman series. Illumination of the neon bulb on signaled trials was followed at the appropriate ISI by shock onset. Light and shock offsets were simultaneous. Intertrial intervals were randomly selected from 10, 15, and 20 sec. Latency of key presses on signaled trials was measured by the klockounter.

Results

Figure 2 presents mean ratings for signaled and unsignaled shock for each ISI. Examination of Figure 2 reveals that unsignaled shocks were consistently rated as more unpleasant than signaled shocks. Furthermore, there was little indication of a systematic effect of interstimulus interval on the ratings. Analysis of variance yielded signaled versus unsignaled shock as the only significant main effect, F (1, 40) = 55.27, $p < .01$. The interaction of sex with signaled versus unsignaled shock was also significant, F (1, 40) = 5.14, $p < .05$. Examination of the data showed that this was because the signaled–unsignaled differences in ratings were considerably greater for females than for males.

Table 1 shows the percentage of key presses that occurred in the .4-sec. interval just prior to the onset of the shock. Accuracy of time estimation was an inverted U-shaped function of ISI, with the most accurate estimation at the .5-sec. ISI. This conclusion was supported by an analysis of variance, which yielded a significant effect of ISI, F (4, 40) = 6.60, $p < .01$, which was composed mainly of a significant quadratic, F (1, 40) = 18.77, $p < .001$, and a significant cubic trend, F (1, 40) = 7.11, $p < .01$. These results essentially replicate those of Teller, Dieter, and Suboski (1972), who used two lights rather than a light and a shock, but the same criterion for "correct" responses.

Table 1 also contains difference scores for the mean ratings on signaled trials. The scores were obtained by subtracting the mean rating for trials on which the key press occurred prior to the end of the shock (anticipatory) from the mean rating for trials on which the key press occurred after shock offset (nonanticipatory). Unfortunately, statistical analysis of these data is not possible because the means are for trials pooled across *S*s within groups. No definition of anticipatory and nonanticipatory responses yields responses in both categories for all *S*s in all groups. Some *S*s in the .25-sec. group made no responses prior to the end of the shock and a number of *S*s in the 2.0- and 4.0-sec. groups made no responses after the end of shock. In addition, analysis is complicated by the fact that the total number of anticipatory and nonanticipatory responses differed widely across ISI groups. About all that can be done is to note that the differences shown in Table 1 are quite large in an absolute sense, for example, relative to the

TABLE 1

ACCURACY OF KEY PRESSES AND NONANTICIPATORY–ANTICIPATORY DIFFERENCES IN AVERSIVENESS RATINGS

Measure	Interstimulus interval (in sec.)				
	.25	.5	1.0	2.0	4.0
% correct key presses	7.6	48.6	44.0	20.6	23.6
Mean rating difference	.629	.908	.828	−.172	−.703

Note.—The first row shows the percentage of key presses that occurred within .4 sec. of the end of the ISI. The second row shows the difference between the mean ratings of trials on which a key press preceded shock offset and nonanticipatory trials.

signaled-unsignaled differences of Fig. 2. The function also appears to be generally of inverted-U form, with anticipatory key press trials rated as considerably less aversive than nonanticipatory trials by the .25-, .5-, and 1.0-sec. groups. Anticipatory trials were rated as more aversive than nonanticipatory by the 4.0-sec. group, but it should be noted that the anticipatory category for the 4.0-sec. group contained many short latency responses.

The results of Exp. I and II were almost completely opposite to each other. However, because of the numerous variations in procedure between the two experiments, the differences in outcomes cannot be attributed to any one factor. Nevertheless, the use of the time-estimation task in Exp. II seems most likely to be the effective variable. Quite possibly *S*s who must estimate the ISI place preparatory responses more accurately than do *S*s who do not time estimate. Thus, lower ratings for signaled than for unsignaled shock would be expected, and the effect of ISI on the ratings of signaled shock would be attenuated or eliminated. Experiment III was designed to investigate the joint effect of the time-estimation task and ISI on ratings of signaled and unsignaled shock.

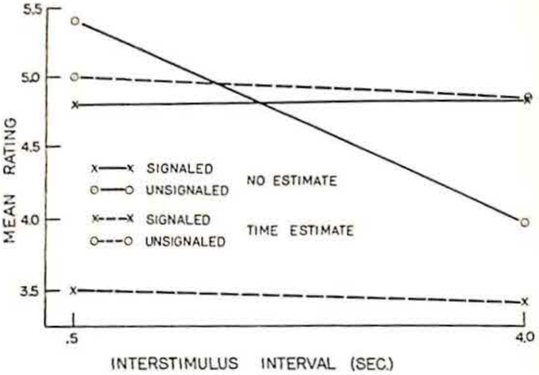


FIG. 3. Mean ratings of signaled and unsignaled shock as a function of ISI for the groups performing and not performing the time-estimation task.

EXPERIMENT III

Method

Procedure.—The *S* population, apparatus, and general procedure were those of Exp. II. Because of the clear-cut outcome, the experiment was terminated when two female and two male *S*s had been randomly assigned to each of the four cells created by factorially combining .5-sec. and 4.0-sec. ISIs and inclusion versus exclusion of the time-estimation task. The instructions were those of Exp. II, but with all mention of the time-estimation task omitted for *S*s not performing the task. All other aspects of the procedure were as described in Exp. II.

Results

Figure 3 presents mean ratings of signaled and unsignaled shock as a function of ISI for the time-estimation and no-time-estimation groups. The *S*s performing the time-estimation task rated signaled shock as considerably less aversive than unsignaled shock at both ISIs. The performance of these *S*s was virtually identical to the performance of the comparable *S*s from Exp. II. The *S*s not time estimating rated signaled shock as less aversive than unsignaled at the .5-sec. ISI and more aversive at the 4.0-sec. ISI. These results closely resemble those of the comparable groups of Exp. 1. Therefore, the difference in results between Exp. I and II was very clearly due to the time-estimation task. Analysis of variance of these data revealed a significant difference between ratings of signaled and unsignaled shock, $F\ (1, 12) = 5.43$, $p < .05$, and a significant Signaled versus Unsignaled Shock × Time Estimation versus No Time Estimation interaction, $F\ (1, 12) = 7.72$, $p < .05$.

In addition to testing the effect of the time-estimation task on shock judgements, the design of this experiment permitted an examination of another variable relevant to the instrumental consequences interpretation. Classical CRs develop over a number of conditioning trials. If CRs occur to a shock UCS because of their instrumental effects in attenuating the noxiousness of the shock, then ratings of signaled shock as less aversive than unsignaled should also develop across trials. However, ratings of shock appear to be most meaningful in terms of relative ratings of signaled and unsignaled shock rather than in absolute terms. Consequently, the inclusion of equal numbers of signaled and unsignaled trials in the present experiment permitted examination of the difference in ratings of signaled and unsignaled shock across trials. Figure 4 displays mean ratings of signaled

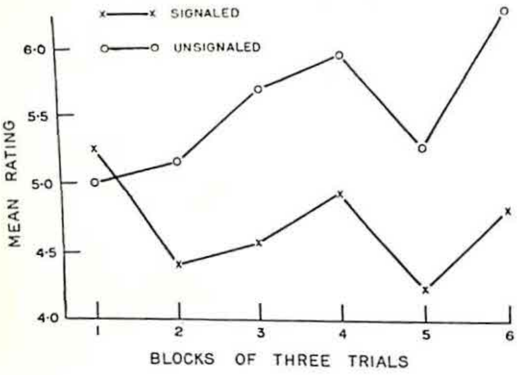


FIG. 4. Mean ratings of signaled and unsignaled shock by the .5-sec., no-time-estimation group as a function of successive blocks of three trials.

and unsignaled shock across blocks of trials for the .5-sec., no-time-estimation group. Of the four groups, this is the one most likely to show an acquisition function. Examination of Fig. 4 shows that the difference between mean ratings of unsignaled and signaled shock changed across trials from a small negative value to large positive values. Analysis of variance of the unsignaled–signaled difference scores yielded a significant effect of trial blocks, $F\ (5, 15) = 13.69,\ p < .01$.

EXPERIMENT IV

The effect of the time-estimation task on the ratings of signaled and unsignaled shock is consistent with the hypothesis that time estimation favors the accurate placement of preparatory CRs that attenuate the noxiousness of the shock. However, performance of the time-estimation task could also have effects for a number of reasons, including diverting *S*'s attention away from the shock, or merely because of the response requirement of pressing the key. Devising appropriate control conditions for all of the possibilities would be difficult, if not impossible. Consequently, in Exp. 4 a completely different approach was adopted.

Prokasy, Ebel, and Thompson (1963) gradually shifted the ISI in eyelid conditioning from a value (.6 sec.) which yields near-optimal levels of conditioning to a distinctly nonoptimal value (2.5 sec.). They found that the percentage of CRs by this group remained equivalent to the high level of a group conditioned at the .6-sec. ISI throughout. These results were interpreted (e.g., Prokasy, 1965) in terms of "shaping" appropriate time estimation in the shifted ISI group. Teller et al. (1972) obtained very similar results by having *S*s estimate the time interval between the onsets of two lights by means of a key press. The ability of *S*s to estimate time accurately at a relatively long ISI was "shaped" by beginning at an ISI favorable to time estimation and gradually increasing the ISI.

The "shaping" results for classical conditioning (Prokasy et al., 1963) and time estimation (Teller et al., 1972) suggest a powerful test of the instrumental consequences hypothesis. Experiments I and III showed that *S*s rate signaled shock as less aversive than unsignaled at short ISIs and more aversive at long ISIs. If the same processes that determine classical conditioning levels and the accuracy of time estimation also determine the ratings of aversiveness of signaled shock, then "shaping" of ratings of signaled shock should be possible. By beginning at a short ISI and making gradual increases in the ISI, signaled shock should continue to be rated as less aversive than unsignaled at long ISIs.

Method

Procedure.—The *S* population, apparatus, and general procedure were as described in Exp. II and III. Seven female and four male *S*s were randomly assigned to each of three groups. Two groups received a constant ISI on all signaled trials: .6 sec. for one group and 2.5 sec. for the other. The third group received a gradual increase in the ISI across blocks of trials, from an initial value of .6 sec. to a final value of 2.5 sec. All *S*s received 80 trials composed of six blocks of 9 signaled trials followed by 2 unsignaled trials and a seventh block of 7 signaled and 7 unsignaled trials. In the final block, signaled and unsignaled trials were ordered by a Gellerman series.

The ISI shifts were made between the trial blocks. The shifted group received ISIs of .6, .8, 1.0, 1.25, 1.6, 2.0, and 2.5 sec. for Trial Blocks 1 through 7, respectively. These values correspond to those used by Prokasy et al. (1963) and Teller et al. (1972), although the present number of trials (61) was considerably fewer than was used in the earlier reports (180). The *S*s were not required to perform the time-estimation task, and the instructions used were those from Exp. III which contained no mention of time estimation.

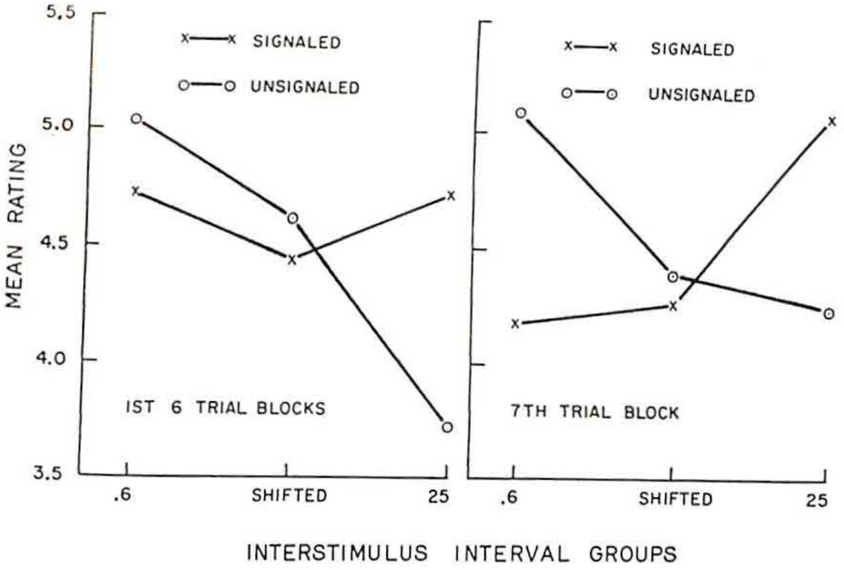


FIG. 5. Mean ratings of signaled and unsignaled shock for each ISI group averaged across the first six trial blocks and for the seventh trial block.

Results

An analysis of variance conducted on the mean ratings of signaled and unsignaled trials for the first six trial blocks yielded no significant effects involving trial blocks. The only significant source of variance was the interaction of ISI groups with signaled versus unsignaled shock, F (2, 30) = 5.85, $p < .01$. This interaction is shown in Fig. 5. Both the .6-sec. and shifted groups rated signaled shock as slightly less aversive than unsignaled, whereas the 2.5-sec. group rated signaled shock as considerably more aversive than unsignaled. The effect of signaled versus unsignaled shock did not differ between the .6-sec. and shifted groups, $F = .14$, but rather was due to the difference between the .6-sec. and shifted groups combined and the 2.5-sec. group, F (1, 30) = 11.56, $p < .01$. Figure 5 also shows the same interaction for the seventh trial block, F (2, 30) = 7.89, $p < .01$, which again was the only significant effect. Again, the effect of signaled versus unsignaled shock did not significantly differ between the .6-sec. and shifted groups, F (1, 30) = 3.11, but the .6-sec. and shifted groups combined did differ from the 2.5-sec. group, F (1, 30) = 12.77, $p < .01$. A post hoc test indicated that the shifted group differed from the 2.5-sec. group, F (1, 30) = 4.90, $p < .05$. Individual group tests revealed that the differences between ratings of signaled and unsignaled shock were significantly different from zero for Groups .6-sec., t (10) = 2.82, $p < .05$, and 2.5-sec., t (10) = 3.60, $p < .05$, but not for the shifted group, $t = .35$.

GENERAL DISCUSSION

The outcomes of these experiments are consistent with the hypothesis that *S*s are capable of making preparatory responses that attenuate the noxiousness of an electric shock. Furthermore, the preparatory responses appear to have many of the characteristics of CRs acquired by classical conditioning. First, the U-shaped function relating intensity ratings of signaled shock to ISI corresponds to the inverted U-shaped function of percentage of CRs with ISI in classical conditioning of skeletal responses. If classical CRs attenuate the aversiveness of a noxious UCS, then the greatest attenuation should occur under conditions which yield the highest conditioning levels. Experiments I and IV demonstrated that this was true for the effects of ISI. Second, conditions which favor the accurate placement of preparatory responses with respect to the UCS should yield an increase in the aversiveness of unsignaled relative to signaled shock. This was found to be the case when *S*s were required to estimate the duration of the ISI by key pressing. In addition, at short ISIs, *S*s rated shock as less aversive on trials when the key press anticipated the shock than on trials when the shock was not anticipated. Little evidence on the effects of time estimation on levels of CRs in classical conditioning seems to be available. However, Coppock (1966) reported improved galvanic skin response (GSR) conditioning at a 5-sec. ISI if *S*s were instructed to count during the ISI. Furthermore, *S*s instructed to key press to the CS show considerably enhanced rates and levels of acquisition of the eyelid CR (Beck, 1939; Miller & Cole, 1936).

Other evidence in favor of the hypothesis that preparatory CRs attenuate the noxiousness of a shock is the finding that ratings of signaled shock as less aversive than unsignaled develop relatively slowly over successive trials. This finding is consistent with the relatively slow development of classically conditioned skeletal responses. The most powerful evidence favoring the preparatory CR hypothesis is the fact that aversiveness ratings can be "shaped" in the same manner as can classical CRs (Prokasy et al., 1963). This finding very strongly indicates that the same processes that determine classical conditioning levels also determine the ratings of aversiveness of signaled shock relative to unsignaled. The accuracy of time estimation can be similarly shaped (Teller et al., 1972). This suggests that the process underlying response "shaping" at long ISIs is the ability of *S* to time preparatory CRs to occur appropriately with respect to the UCS, as Prokasy (1965) has argued.

The results of the present experiments are consistent with the majority of the previous investigations of signaled and unsignaled shock in humans. The effects of varying ISI were completely consistent with the indications from Kimmel's (1967) results that signaled shock is rated as less intense than unsignaled at short ISIs and as more intense at long ISIs. However, there are some inconsistencies between the present results and earlier reports. The present findings would indicate that at relatively long ISIs unsignaled shock should always be rated as less aversive than, and preferred over, signaled shock. Such findings have sometimes been reported (e.g., Furedy & Doob, 1971a, 1971b; Kimmel, 1967), but a number of investigators have found nonsignificant differences (Furedy, 1970; Furedy & Chan, 1971) and even a significant preference for signaled shock at ISIs of 4 sec. or greater (Badia et al., 1967; Lanzetta & Driscoll, 1966; Pervin, 1963). In addition, Pervin reported that signaled shocks at a 5-sec. ISI were rated as less painful than unsignaled shocks.

The present experiments have been interpreted as supporting the hypothesis that classical CRs are acquired because of their instrumental effects in modifying the consequences of the UCS. Nevertheless, all of the evidence of the present studies is indirect. No attempt was made to measure or examine the hypothesized CRs. It is possible that *S*s make preparatory responses which attenuate the aversiveness of shocks and that these responses have no relationship to any of the CRs commonly conditioned with a shock UCS. That is, some of the CRs acquired in classical conditioning may be acquired because of their instrumental consequences, whereas other CRs may not require instrumental consequences. This possibility could account for reported failures to find any relationship between magnitudes of autonomic CRs and ratings of shock intensity (Furedy, 1970; Furedy & Doob, 1971a, 1971b), suggesting that the CRs either were acquired in the apparent absence of an instrumental effect of the CR on the shock UCS or they were not, in fact, CRs at all. If the former is the case, then the present evidence indicative of instrumental effects of CRs with a shock UCS merely serves as a reminder that subtle instrumental processes may, and undoubtedly do, occur in classical conditioning.

The possibility that classical CRs were not acquired in the experiments by Furedy and his associates deserves greater consideration than has so far been given. In several studies (Furedy, 1970; Furedy & Doob, 1971a), a trace conditioning procedure with a .3-sec. CS and a 5-sec. ISI was used with 15 conditioning trials. The *S*s also received 15 unreinforced CSs (CS−) and 15 unpaired UCSs interspersed among the reinforced (CS+) trials. Other experiments used similar parameters, except that in one experiment (Furedy & Doob, 1971a, Exp. III) delay conditioning was used and another (Furedy & Doob, 1971b) used delay conditioning and an 8-sec. ISI. These parameters appear distinctly unfavorable to the acquisition of CRs, and the evidence that conditioning was obtained is considerably less than convincing. The primary evidence of conditioning was significant differences in the magnitude of anticipatory responses to CS+ (light or tone) and CS− (tone or light). Nevertheless, these measures probably reflect sensitization rather than conditioning. Response magnitude to CS+ (Furedy, 1970; Furedy & Doob, 1971b) and the difference in response magnitude to CS+ and CS− (Furedy, 1970) was greatest on the first block of three trials and decreased across trials thereafter. These changes are diametrically opposed to those expected as a consequence of the acquisition of CRs.

The measures in the Furedy (1970) and Furedy and Doob (1971a, 1971b) studies were autonomic responses, the GSR, and volume-pulse change. The latter is a measure of peripheral vasoconstriction that so far has not been widely recorded. However, the extent to which GSR conditioning reflects the operation of classical conditioning processes has been the subject of extended debate and investigation (cf. Dengerink & Taylor, 1971). Many investigators of GSR conditioning differentiate between short latency responses, which presumably do not reflect the classical conditioning process, and longer latency responses, which presumably are the result of associative conditioning. Furedy and his associates have not made such a distinction, but rather have considered all responses with latencies greater than 1 sec. as CRs. Thus, the failure of Furedy and his associates to obtain any relationship between CR magnitudes and ratings of intensity may well be attributable to the failure to use parameters that yielded anticipatory CRs.

REFERENCES

Badia, P., Suter, S., & Lewis, P. Preference for warned shock: Information and/or preparation. *Psychological Reports*, 1967, **20**, 271–274.

Beck, L. H. Conditioning and the coordination of movements. *Journal of General Psychology*, 1939, **20**, 375–397.

Coppock, H. W. GSR conditioning at a long CS–US interval mediated by *S*'s counting behavior. *Psychonomic Science*, 1966, **4**, 155–156.

Dengerink, H. A., & Taylor, S. P. Multiple responses with differential properties in delayed galvanic skin response conditioning: A review. *Psychophysiology*, 1971, **8**, 348–360.

Furedy, J. J. Test of the preparatory-adaptive-response interpretation of aversive classical autonomic conditioning. *Journal of Experimental Psychology*, 1970, **84**, 301–307.

Furedy, J. J., & Chan, R. M. Failures of information to reduce rated aversiveness of unmodifiable shock. *Australian Journal of Psychology*, 1971, **23**, 85–94.

Furedy, J. J., & Doob, A. N. Autonomic responses and verbal reports in further tests of the preparatory-adaptive-response interpretation of reinforcement. *Journal of Experimental Psychology*, 1971, **89**, 258–264. (a)

Furedy, J. J., & Doob, A. N. Classical aversive conditioning of human digital volume-pulse change and tests of the preparatory-adaptive-response interpretation of reinforcement. *Journal of Experimental Psychology*, 1971, **89**, 403–407. (b)

Gormezano, I., & Fernald, C. D. Human eyelid conditioning with paraorbital shock as the US. *Psychonomic Science*, 1971, **25**, 88–90.

Gormezano, I., & Moore, J. W. Classical conditioning. In M. H. Marx (Ed.), *Learning: Processes*. New York: Macmillan, 1969.

Kimmel, E. Judgments of UCS intensity and diminution of the UCR in classical GSR conditioning. *Journal of Experimental Psychology*, 1967, **73**, 532–543.

Lanzetta, J. T., & Driscoll, J. M. Preference for information about an uncertain but unavoidable outcome. *Journal of Personality and Social Psychology*, 1966, **3**, 36–102.

Miller, J., & Cole, L. E. The influence of a "voluntary" reaction upon the development and the extinction of the conditioned eyelid reaction. *Journal of Genetic Psychology*, 1936, **48**, 405–440.

Perkins, C. C., Jr. An analysis of the concept of reinforcement. *Psychological Review*, 1968, **75**, 155–172.

Perkins, C. C., Jr., Seymann, R. G., Levis, D. J., & Spencer, H. R., Jr. Factors affecting preference for signal-shock over shock-signal. *Journal of Experimental Psychology*, 1966, **72**, 190–196.

Pervin, L. A. The need to predict and control under conditions of threat. *Journal of Personality*, 1963, **31**, 570–587.

Prokasy, W. F. Classical eyelid conditioning: Experimenter operations, task demands, and response shaping. In W. F. Prokasy (Ed.), *Classical conditioning*. New York: Appleton-Century-Crofts, 1965.

Prokasy, W. F., Ebel, H. C., & Thompson, D. D. Response shaping at long interstimulus intervals in classical eyelid conditioning. *Journal of Experimental Psychology*, 1963, **66**, 138–141.

Teller, K. J., Dieter, R., & Suboski, M. D. Time estimation and the interstimulus interval function in classical conditioning. *Journal of Experimental Psychology*, 1972, **95**, 445–447.

Tursky, B., Watson, P. D., & O'Connell, D. N. A concentric shock electrode for pain stimulation. *Psychophysiology*, **1**, 1965, 296–298.

(Received June 19, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 416–424

EXPERIMENTAL MANIPULATION OF ORIENTING REFLEX DURING SEMANTIC CONDITIONING[1]

ANN K. LODWIG[2]

University of California, Los Angeles

An orienting reflex evoked during conditioning is assumed to enhance perception and augment the magnitude of the subsequent conditioned responses (CRs). This hypothesis was tested by semantically conditioning the galvanic skin response (GSR) using two CS–UCS pairs concurrently: CS_1–noise, CS_2–tone. The tone was a signal for a simple reaction time (RT) measure. Two groups of 64 university students were *S*s. Twenty conditioning trials were embedded in a list of filler words. The Nonreversal group received the same two CS–UCS pairs for all 20 conditioning trials. An orienting reflex (OR) was evoked in the Reversal group when the critical stimulus contingencies were reversed, so that for the last 10 conditioning trials the particular CS–UCS pair was reversed from what it had been during the first ten trials. All *S*s manifested reliable semantic conditioning and generalization, with High-OR *S*s giving significantly larger mean CRs than Low-OR *S*s. Reversal group *S*s showed augmented CRs, better differential conditioning, and slower RTs following the reversal, supporting the initial hypothesis. The results are discussed in terms of the role of the OR and arousal-raising treatments in conditioning.

Individual differences in the magnitude of the orienting reflex (OR) have been found to be reliably related to differences in the semantic conditioning of autonomic responses and to a number of other complex behaviors (Brotsky, 1968; Maltzman, 1967, 1968; Maltzman & Mandell, 1968). In these studies, *S*s categorized as high orienters on the basis of the magnitude of their response to the first presentation of a neutral word or of the unconditioned stimulus (UCS) showed better differential conditioning and better generalization than Low-OR *S*s. These results demonstrated the close relationship between the level of OR and the strength of semantic conditioning. This relationship suggests the possibility that manipulation of the OR during the acquisition of the CR could strengthen conditioning.

Sokolov (1963) has argued that the role of the OR in conditioning is facilitation of the *S*'s perception of his environment. Experimental evidence given for this contention is the increase in sensitivity of the auditory and visual systems when a stimulus in a different modality evoked an OR. In view of (*a*) the experimental results which indicate that the level of OR is related to the strength of conditioning and (*b*) the hypothesized role of the OR in facilitating perception, evoking an OR during semantic conditioning should enhance perception of the relevant contingencies and should strengthen differential conditioning. The experiment reported here was designed to test this proposition.

---

[1] This study was carried out as a dissertation project submitted in partial fulfillment of the requirements for the PhD degree at the University of California, Los Angeles. I wish to thank Irving Maltzman for his unstinting help and supervision, and Harry Jerison for a critical reading of this manuscript. My appreciation also goes to Barry Langdon, Linn Ng, and Delores Adams for their help during the analysis of the data. The major portion of this research was carried out during the tenure of a predoctoral Research Fellowship from the National Institutes of Mental Health, United States Public Health Service Award MH-17839. Additional support was obtained from United States Public Health Service Grant MH-04684, Irving Maltzman, principal investigator. Computing assistance was obtained from the Health Sciences Computing Facility at the University of California, Los Angeles, and was sponsored by National Institutes of Health Grant FR-3 and by National Institutes of Health Special Resources Grant RR-3. Some of the results of this study were reported at the Western Psychological Association meetings, Los Angeles, April 1970.

[2] Requests for reprints should be addressed to Ann K. Lodwig, Department of Psychiatry, NPI-MR, 760 Westwood Plaza, Los Angeles, California 90024.

METHOD

Subjects.—The *Ss* were 128 introductory psychology students at the University of California, Los Angeles. These *Ss* were divided equally into two treatment groups, a Reversal group (Group R) and a Nonreversal group (Group NR). An equal number of male and female *Ss* was run in each group. Participation in the experiment was voluntary, but this service partially fulfilled a course requirement.

Apparatus.—A two-room experimental suite with a one-way observation mirror was used for running this experiment. In the observation room, *E* operated the stimulus presentation and recording equipment. The stimuli were recorded on magnetic tape and played back to *S* by a Webcor stereophonic tape deck amplified by a Scott stereophonic amplifier. The *Ss* listened to the stimuli through a Calrad Stereophonic headset. A General Radio random noise generator was used for producing the white noise, and a General Radio R-C oscillator was used to produce the 1,000-Hz tone. The *Ss* heard the noise at 110 db., the tones at 70 db., and the words at 70–75 db. All stimulus events, i.e., each word, tone, and burst of noise, were recorded automatically on the chart of a Sanborn Model 320 DC amplifier-recorder.

The *Ss* were seated comfortably in an electrically and acoustically shielded room on the other side of the one-way mirror. The *S*'s galvanic skin response (GSR) was measured by circular zinc electrodes 20 mm. in diameter embedded in 5-mm.-deep plastic cups. The cups were filled with electrode paste made from agar and a 1% zinc sulphate solution. Plastic pressure clips were used to attach the electrodes to the *S*'s palms. Skin resistance was measured continuously during the experiment by a Darrow type bridge supplying a 40-μa. constant current. The GSR data were fed directly into the Sanborn amplifier-recorder, with the paper speed at 1 mm/sec.[3]

The *Ss* were provided with a foot pedal on which they comfortably rested their right foot. A Hunter Klockounter was used to record reaction time (RT) between the presentations of the tone and lifting the foot off the pedal.

Materials and procedure.—In order to evoke an OR during conditioning for the experimental *Ss*, it was necessary to introduce some kind of change in ongoing stimulation. A dual-conditioning procedure was used that provided a change relevant to the conditioning procedure itself. Semantic conditioning of the GSR to one CS word was established by presenting a number of trials with the CS followed by a loud noise as the UCS. Simultaneous conditioning of the GSR to a second CS word was accomplished by instructing *S* to lift his foot off a pedal as soon as he heard a tone, which followed the second CS. In order to evoke an OR during conditioning, the signal (tone) for the pedal release response was switched to the CS word which had preceded the loud noise, while the UCS (noise) was paired with the CS which previously preceded the tone.

All of the data for a particular *S* were collected at a single session which lasted 50 min. Initially, all *Ss* were presented with a list of 20 words to habituate the GSR to the occurrence of words prior to the first conditioning trial. Immediately following the habituation list, *Ss* heard the 20 conditioning trials embedded in a list of 57 neutral filler words. The tape concluded with 20 additional words of which 6 were test words for semantic generalization. Throughout the session, the interword interval varied from 9 to 15 sec. During conditioning and generalization, between two and five filler words separated the CSs or the generalization test words.

During the conditioning phase, two CS words each occurred in 10 conditioning trials in the following sequence: CS_1, CS_1, CS_2, CS_2, CS_2, CS_1, CS_1, CS_1, CS_2, CS_2, CS_1, CS_1, CS_2, CS_2, CS_2, CS_1, CS_1, CS_1, CS_2, CS_2. A 1-sec. 1,000-Hz. tone at 70 db., or a 1-sec. burst of 110-db. white noise occurred 5.5 sec. after the CS word. Within each group, half of the *Ss* had STEM as CS_1, and CARPET as CS_2, and these were reversed for the other half of the *Ss* in each group.

For the 64 *Ss* in Group NR, CS_1 was always paired with the tone or the burst of white noise, and CS_2 was always paired with the burst of white noise or the tone, respectively. For the 64 *Ss* in Group R, the CS–UCS contingencies were reversed halfway through the conditioning trials. After five trials with CS_1–tone and five trials with CS_2–noise, *Ss* were switched without warning to CS_1–noise and CS_2–tone.

The two generalization test words were FLOWER and RUG. The CS and generalization test words were chosen from the Minnesota revision of the Kent-Rosanoff norms (Russell & Jenkins, 1954) because of the high relative frequency of occurrence of the CS words and because the most commonly given associate to these words occurred with nearly the same frequency. Of 1,008 *Ss* in the Russell and Jenkins study, 460 gave the word RUG in response to CARPET and 402 gave the response FLOWER to STEM. The filler words were selected from the revised Kent-Rosanoff norms so that none of them had either of the CS words or generalization test words in their response hierarchies.

The *Ss* were assigned without bias to one of the treatment groups and to one of the word order subgroups. After *Ss* were seated, *E* attached the electrodes and read the instructions which informed *Ss* that they would hear words, tones, and bursts of noise with intervals of silence between them. The *Ss* were asked to release the pedal as soon as they heard the tone, but otherwise to sit quietly and relax. The *E* then placed the earphones on *S*, left the room,

---

[3] A Starling Corporation photoplethysmograph was attached with Scotch electrical tape to the forehead of *S* in the region of the left temporal artery and secured with a Velcro headband. Analysis of the vasomotor data revealed a large portion of it to be highly unreliable, and no further reference to it will be made in this report.

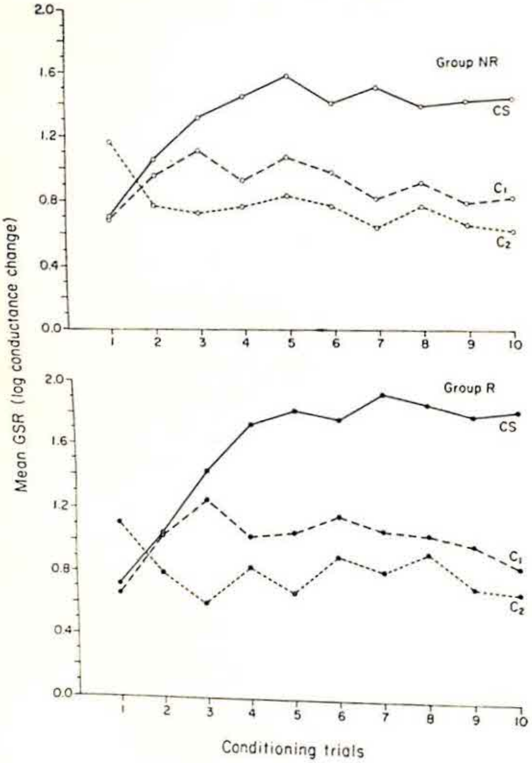


FIG. 1. Mean GSRs of the Nonreversal and Reversal *S*s to the two CSs and the control words. (Data from CS_1 and CS_2, and their associated control words, are combined for each group, so that each trial represents the average of the *S*'s performance on that presentation of CS_1 and CS_2.)

and started the stimulus tape. There were 20 sec. of silence on the tape before the first word. The experiment then continued as programmed on the tape recorder for 35 min. Immediately following the generalization list, the electrodes were removed and *S*s were administered a postexperimental questionnaire to determine whether they had discovered the CS–UCS and the CS–generalization test word contingencies. The questionnaire included ratings of the CS words, the generalization words, and neutral filler words on an evaluative scale of the semantic differential (Osgood, Suci, & Tannenbaum, 1957). The *S*s were also administered the Pittsburgh revision of the Taylor Manifest Anxiety Scale (TMAS) (Bendig, 1956).

The GSR was defined as the maximum decrease in skin resistance which occurred in the .5- to 5.0-sec. interval following the stimulus onset. A log conductance-change transformation was performed on the data prior to analysis.

Each *S*'s OR was determined by his GSR to the first word in the habituation list. Following Maltzman's (1967) procedure, *S*s in each treatment group responding above the median magnitude of the distribution of these responses were classified as High-OR *S*s, while *S*s responding below the median were classified as Low-OR *S*s.

RESULTS

Individual and group differences were evaluated by mixed-design analyses of variance and analyses of covariance. The analyses examined the following main effects, where applicable: (*a*) the between-*S* variables of OR (high vs. low), sex, UCS order (whether the noise or the tone was paired with the first CS), CS order (whether CARPET or STEM occurred as the first CS), and groups, (*b*) the within-*S* variables of trials, words, or the CS and its immediately preceding (C_1) and following (C_2) control words, and trial blocks (prereversal vs. postreversal). All of the effects that are reported were significant beyond the .05 level.

Habituation

There was a monotonic decline in the magnitude of the GSRs evoked by the different words in the habituation list for all of the *S*s. The *S*s classified as high orienters on the basis of the magnitude of their GSR to the initial word in the list maintained this greater response magnitude as compared to the Low-OR *S*s throughout the habituation session, *F* (1, 96) = 58.27.

A probable sampling bias was indicated by the initial performance in the habituation phase of the experiment. Group R had a reliably larger mean GSR to the first word in habituation than Group NR, *F* (1, 96) = 4.56. The difference between groups was never found to reach the .05 level of significance during subsequent testing until the effect of the reversal of stimulus contingencies for Group R was examined. The reversal apparently augmented the difference between these groups.

Conditioning

A within-*S* criterion of differential conditioning was considered met when the statistical analysis of the conditioning data showed (*a*) a significant effect for words, indicating that the GSRs to the CS were reliably larger than to the control words over all conditioning trials, and (*b*) a significant Words × Trials interaction in which the responses to the CS showed a greater positive slope than the responses to the control words.

Evidence of differential semantic conditioning for Groups R and NR is presented in Fig. 1, in which the change in GSR to the CS is compared to the change in GSR to the control words immediately preceding (C_1) and following (C_2) the CS, regardless of which CS was presented and whether the CS was paired with the noise or with the tone-pedal response. The statistical analysis of Group NR revealed that with CARPET as the CS, the GSRs to the CS were significantly greater than to the control words, F (2, 96) = 76.41, and there was a significant Words × Trials interaction, F (18, 864) = 7.49. Likewise, with STEM as the CS, words was a significant effect, F (2, 96) = 46.91, and there was also a significant Words × Trials interaction, F (18, 864) = 6.20. Analogous statistical evidence for the reliability of differential conditioning in Group R was not sought, since such an analysis appeared unnecessary after it was done for Group NR. Analyses comparing the effectiveness of the two CS words in Groups NR and R indicated that they did not differ reliably, and so they were combined for Fig. 1 and for subsequent analyses.

Reversal of contingencies.—Prior to the reversal of CS-UCS contingencies, there was no significant difference in conditioned response (CR) magnitude between Groups R and NR. Following the reversal, CRs of Group R were reliably larger than CRs of Group NR, F (1, 120) = 7.21. In order to determine the reliability of the apparent difference between Groups R and NR prior to the reversal, an additional analysis was conducted on the results from the fifth conditioning trial. The results of this analysis failed to reach statistical significance.

Assessment of group differences in differential conditioning controlled for the inequality of *Ss'* OR magnitude in each group noted in the presentation of the habituation results. Group differences were examined by analyses of covariance which used each *S*'s OR as the covariate for the analysis of the prereversal trials. The level of prereversal conditioning, each *S*'s mean CR on the prereversal trials, was the covariate for assessment of group differences on the postreversal block of trials. The covariance analyses compared three measures of the performance of Groups R and NR on the two blocks of trials. The three measures were *Ss'* mean CRs, the mean difference between CRs and GSRs to C_1, and the mean difference between CRs and GSRs to C_2.

For the prereversal block of trials, with each *S*'s OR as the covariate, the two groups did not differ on any of the three measures. This analysis validated the assumption that with correction for the sampling bias, the two groups showed no reliable differences in differential conditioning prior to the reversal of stimulus contingencies for Group R.

The covariance analysis of the postreversal block of trials was designed to control for the possibility that the observed difference between Groups R and NR following the reversal was a result of differences in the level of conditioning achieved by *Ss* in the two groups during the first half of the conditioning trials. Two analyses of covariance were performed. The first used as covariates *S*'s OR and his mean CR on the prereversal trials. The second analysis used as the covariate only *Ss'* mean CRs on the prereversal trials. The results of both analyses were essentially identical. When the covariate of mean prereversal CR was included in the analysis, the value for *S*'s OR contributed very little to the outcome. Consequently, only the results of the analysis using the mean prereversal CR as a covariate are reported in detail.

Comparing Groups NR and R with respect to their adjusted postreversal mean CRs (the dependent variable), Group R gave larger responses than Group NR, F (1, 125) = 11.11. Further, within each group, the postreversal mean CR was dependent upon the prereversal mean CR (the covariate), justifying the covariance

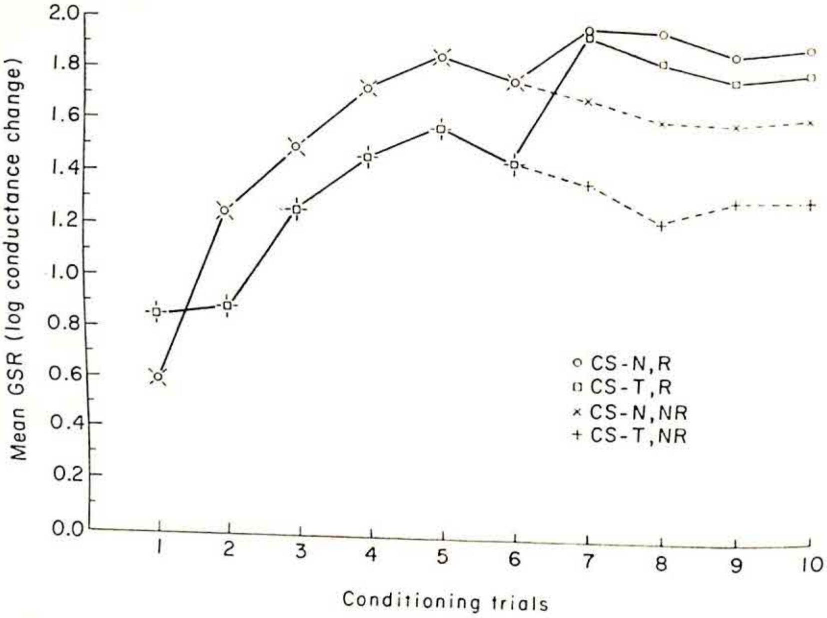


FIG. 2. Mean CRs, as a function of the UCS, of the Nonreversal and Reversal *S*s. (Data from CS_1, and CS_2 are combined, and Groups R and NR are combined prior to the reversal of stimulus contingencies on trial 6 for Group R.)

analysis, *F* (1, 125) = 105.64. The slope of the regression line of the dependent variable on the covariate was statistically equal in both groups.

With respect to the postreversal mean difference between CRs and GSRs to the control words (C_1 and C_2), Group R showed reliably larger differences, i.e., better differential conditioning, than did Group NR, *F* (1, 125) = 6.23, for CR–GSR to C_1, and *F* (1, 125) = 4.36, for CR–GSR to C_2. Again, in both cases the values obtained for these dependent variables (CR–GSR to C_1, CR–GSR to C_2) were dependent upon the covariate, *F* (1, 125) = 12.16, for CR–GSR to C_1, and *F* (1, 125) = 17.69, for CR–GSR to C_2. Also, in each case, the slope of the regression line of the dependent variable on the covariate was statistically equal for both Groups R and NR.

Analysis of variance revealed that within each group, there was no overlap in average performance between High-OR and Low-OR *S*s. Table 1 gives the trial means on the fifth and seventh conditioning trials for the High-OR and Low-OR *S*s in the two groups. The reversal of contingencies for Low-OR *S*s in Group R had the effect of elevating their CR performance to the level of the High-OR *S*s in Group NR.

Nature of the UCS.—The previous analyses ignored the nature of the UCS, noise, or tone-pedal response and examined the effects of reversal versus nonreversal in terms of the responses to the CS words averaged over the UCS contingencies. The following analyses of variance examined the CS–UCS contingencies in terms of their effects of the CRs in the two groups.

Prior to the reversal, Groups R and NR did not differ reliably in terms of the magnitude of CRs induced by the noise and by

TABLE 1

MEAN CONDITIONED RESPONSES (CRS) OF LOW-ORIENTING REFLEX (OR) *S*S AND HIGH-OR *S*S IN GROUPS NR AND R ON TRIAL 5 AND TRIAL 7

OR *S*s and group	Trial 5	Trial 7
Low		
NR	1.31	1.27
R	1.65	1.71
High		
NR	1.84	1.75
R	1.97	2.12

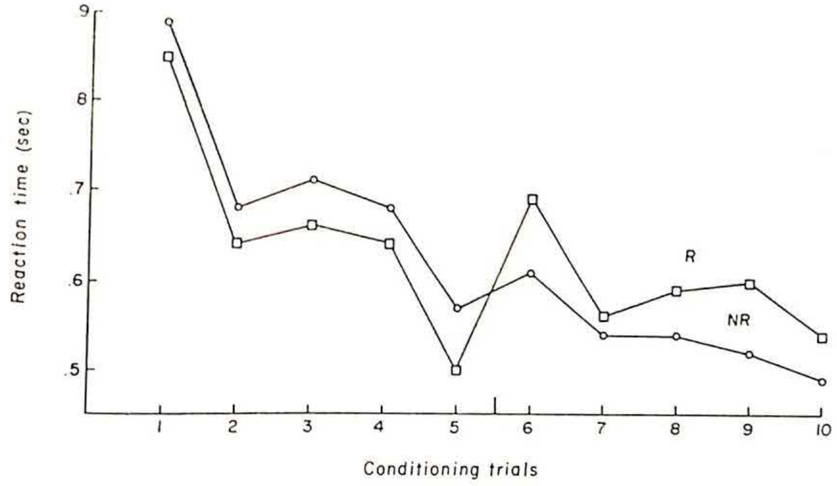


FIG. 3. Mean RTs of the Nonreversal and Reversal groups.

the tone. As expected, the CRs in the noise condition were reliably greater than the CRs evoked in the tone condition, regardless of group, F (1, 122) = 18.93.

Analyses of the postreversal trials showed that there was a significant difference between groups, F (1, 112) = 10.95. The CRs produced by Group R were reliably greater than the CRs produced by Group NR. The type of UCS also produced a significant effect, F (1, 112) = 19.41, and interacted with groups and trials, F (4, 448) = 2.66. Figure 2 shows the nature of the interaction between UCS, groups, and trials, as well as the overall changes in CRs before and after the reversal in CS–UCS contingencies. The reversal in Group R seemed to elevate performance on both CS–noise and CS–tone trials, diminishing the difference between them in the last block of trials in comparison with Group NR.

Reaction time.—Aside from the expected results, for example, that RT would decrease with trials, the finding of major interest was that the reversal of CS–UCS contingencies produced a highly reliable disruptive effect upon the RT of Group R as compared to Group NR, F (1, 96) = 13.36, for the Groups × Trial Blocks interaction. Figure 3 shows that for the first five trial pairs, Group R was somewhat superior to Group NR. Following the reversal in CS–UCS contingencies for Group R, they showed a marked increase in RT which still has not returned to the same level of performance displayed by Group NR by the end of the conditioning phase of the experiment.

Generalization

Generalization of the GSR.—The criterion of generalization was similar to that for conditioning. The magnitude of the GSR evoked by the generalization test word (GS) was compared to magnitude of the responses evoked by the immediately preceding and following neutral control words, C_1 and C_2, respectively.

Reliable differential semantic generalization was obtained in both Groups R and NR. The analyses of variance revealed words as a significant effect, F (2, 64) = 8.39, for Group R, and F (2, 64) = 5.81, for Group NR, with *S*s giving a larger response to the generalization test words than to the control words. There was no reliable difference in generalization between Groups R and NR.

Analysis of the generalization data revealed that level of OR was a significant variable for both groups, with High-OR *S*s giving higher mean responses than Low-OR *S*s, F (1, 32) = 14.93, for Group R and F (1, 32) = 8.79, for Group NR.

Semantic differential ratings.—Following the experiment, *S*s were asked to rate on

TABLE 2

MEAN SEMANTIC DIFFERENTIAL RATINGS GIVEN THE CONTROL WORDS, CONDITIONED STIMULI (CSs), AND GENERALIZATION STIMULI (GSs) BY GROUP NR

CS–UCS contingency	Word						
	Control	CARPET	RUG	FLOOR	STEM	FLOWER	PLANT
CARPET-N, STEM-T	4.50	2.16	2.78	4.03	4.78	4.91	4.31
CARPET-T, STEM-N	4.31	4.78	4.28	4.06	2.13	4.13	3.81

the pleasant–unpleasant evaluative scale of the semantic differential four control filler words, the two CS words, the two GS words, and two words related to the CSs but not heard during the experiment. Values of from 1 to 7 were given the words on the basis of whether *S* found them "very unpleasant" (a value of 1), "neutral" (4), or "very pleasant" (7). An *S*'s ratings of the four control filler words were averaged to obtain a single rating for them from each *S*.

Analyses of variance were performed on the semantic differential data from Group NR. Group R was excluded because the reversal of CS–UCS contingencies led many *S*s to give two ratings to each of the CS words and to give ambiguous ratings to the generalization test words.

Table 2 shows the mean ratings given words by *S*s in Group NR. The differences among ratings of the different words were significant, *F* (6, 336) = 11.81. The analysis also yielded a significant effect for CS–UCS contingency, *F* (6, 336) = 42.04. The semantic differential ratings thus showed evidence of conditioning and semantic generalization which varied as a function of the type UCS presented during conditioning.

The finding that the actual generalization test words reflected the conditioning procedure more closely than the two related words not heard during the experiment suggests that the repetition of the generalization test words may have had two possible effects: (*a*) It heightened the discrimination between the two kinds of words, and (*b*) it increased the level of semantic generalization to the two repeated words used in the generalization test. Confirming evidence for the latter effect was found in the GSR generalization data, in which maximum generalization of the GSR occurred on the second or third presentation of the test word. Repetition of the CS during conditioning was one of its distinguishing features, and it is likely that repetition of the generalization test word led *S*s to anticipate that it, like the CS, would be followed by the UCS.

The OR and Anxiety

Level of OR did not correlate with the TMAS scores. This confirms the findings of other studies which indicate that the GSR, used appropriately as a measure of the OR, is not simply a redundant measure of anxiety (Belloni, 1964; Maltzman, 1967).

DISCUSSION

This study demonstrated that by evoking an OR during conditioning, subsequent differential conditioning performance was improved. This strengthening in conditioning occurred upon the reversal of stimulus contingencies for Group R *S*s and was manifested by augmented responses to the CS but not to the control stimuli. Further evidence for the effectiveness of the reversal in evoking an OR was that the postreversal RTs of Group R *S*s were significantly slower than those of *S*s in Group NR.

Thus in the present study, both summation and disruptive effects of evoking an OR were demonstrated simultaneously. The first of these two ways in which the occurrence of an OR interacts with ongoing activity, strengthening of the CR by summation with the OR, is characterized by Anokhin (1965) as assimilation of the OR by the ongoing activity of the organism. According to his description, summation of the excitation from the OR with the dominant activity of the organism occurs in such a way that the OR contributes to the dominant condition that precedes it. In the present experiment, strengthening of the CR following the occurrence of an OR was a discrete event, in that the increment in GSR responding did not occur to the control words. This suggests that the mechanism by which the OR influences conditioning can be quite specific and does not necessarily elevate general autonomic nervous system reactivity.

Sokolov (1965) suggests that this kind of summation effect is manifested when an OR occurs after the prior repeated use of a uniform CS has already led to some inhibition of the CR. At asymptotic CR performance, continuing the presentation of the CS leads to a gradual decline in the level of the CR, and the introduction of a discriminative stimulus strengthens the CR. Sokolov attributes this effect to the activating influence of the OR on the reticular system.

In recent studies of the effect of postpeak acquisition trials on extinction of the conditioned GSR (Bishop & Kimmel, 1969; Lanning & Yaremko, 1971; Schramm & Kimmel, 1970), the authors hypothesized that postpeak acquisition trials increase response inhibition, which is the reason trials to extinction vary inversely with the number of postpeak trials. Evocation of an OR during conditioning, relevant to the conditioning contingencies, provides an additional critical variable for studying the interaction between excitatory and inhibitory processes during response acquisition.

The disruptive effect of the midsession OR on RT was anticipated and has been described by Sokolov (1963). Anokhin (1965) characterizes this form of interaction between the OR and other activities of the organism as conflict brought about by the OR which can, depending on the strength of the excitatory processes that comprise it, inhibit all other acts. In the present study, both summation and the disruptive effect of evoking an OR were demonstrated simultaneously, verifying the fact that the effects of an OR differ with respect to the particular response system under investigation. Not only do these differences occur when comparing ongoing motor activity with responses involving the autonomic system, but there are important individual differences in characteristic autonomic response patterns. This has been a common finding and is discussed at some length by Lacey and Lacey (1958) and by Maltzman and Mandell (1968).

There were two main findings of the present experiment pertinent to the effect of an increase in arousal on differential conditioning. The first was the finding that within Groups R and NR, the noise UCS resulted in higher magnitude CRs than the tone UCS. The second finding relevant to the issue was that the Reversal *S*s maintained a higher level of differential conditioning during the last trials as compared to the Nonreversal *S*s.

We can consider that in the present experiment, tonic arousal level was held constant by giving the same number of both noise and tone trials to all of the *S*s. The phasic ORs, more intense to the noise than to the tone trials, led to the establishment of the CRs and were responsible for the difference in amplitude of the CRs resulting from the noise and tone trials. In this sense, the phasic OR is a determiner of learning as a function of its role in producing changes in the CNS conducive to improved discrimination and performance.

This role of the phasic OR as a determiner of learning can account for the second finding pertinent to the arousal issue: improved differential conditioning performance of the Reversal *S*s as compared to the Nonreversal *S*s during the last conditioning trials. The OR evoked by the reversal and by the consequent increase in uncertainty about the conditioning situation itself increased the ability of *S*s to discriminate critical stimuli and provided a brief, appropriate increase in arousal that helped maintain the CR at a higher level than was the case for the Nonreversal *S*s. Such an interpretation relates the results of this semantic conditioning experiment to the extensive literature on arousal and performance (Berlyne, 1960), in particular by its demonstration that conditioning is enhanced by arousal-raising treatments.

REFERENCES

ANOKHIN, P. K. The role of the orienting-exploratory reaction in the formation of the conditioned reflex. In L. G. Voronin, A. R. LURIA, E. N. SOKOLOV, & O. S. VINOGRADOVA (Eds.), *Orienting reflex and exploratory behavior.* Washington, D. C.: American Institute of Biological Sciences, 1965.

BELLONI, M. L. The relationship of the orienting reaction and manifest anxiety to paired-associates learning. Unpublished doctoral dissertation, University of California, Los Angeles, 1964.

BENDIG, A. W. The development of a short form of the MAS. *Journal of Consulting Psychology,* 1956, **20**, 384.

BERLYNE, D. E. *Conflict, arousal, and curiosity.* New York: McGraw-Hill, 1960.

BISHOP, P. D., & KIMMEL, H. D. Retention of habituation and conditioning. *Journal of Experimental Psychology,* 1969, **81**, 317–321.

Brotsky, S. J. Classical conditioning of the galvanic skin response to verbal concepts. *Journal of Experimental Psychology*, 1968, **76**, 244–253.

Lacey, J. I., & Lacey, B. C. Verification and extension of the principle of autonomic response stereotypy. *American Journal of Psychology*, 1958, **71**, 50–73.

Lanning, A. W., & Yaremko, R. M. Resistance to extinction in GSR conditioning: Effects of postpeak CR training and preextinction. *Journal of Experimental Psychology*, 1971, **88**, 433–435.

Maltzman, I. Individual differences in "attention": The orienting reflex. In R. M. Gagne (Ed.), *Learning and individual differences*. Columbus, Ohio: Charles E. Merrill Books, 1967.

Maltzman, I. Theoretical conceptions of semantic conditioning and generalization. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Maltzman, I., & Mandell, M. P. The orienting reflex as a predictor of learning and performance. *Journal of Experimental Research in Personality*, 1968, **3**, 99–106.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.

Russell, W. A., & Jenkins, J. J. *The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test*. (Tech. Rep. No. 11), Contract N80NR-66216, Minneapolis: University of Minnesota, 1954.

Schramm, C. F., & Kimmel, H. D. Resistance to extinction in GSR conditioning following different numbers of postpeak acquisition trials. *Journal of Experimental Psychology*, 1970, **84**, 239–243.

Sokolov, E. N. *Perception and the conditioned reflex*. New York: Macmillan, 1963.

Sokolov, E. N. The orienting reflex, its structure and mechanisms. In L. G. Voronin, A. R. Luria, E. N. Sokolov, & O. S. Vinogradova (Eds.), *Orienting reflex and exploratory behavior*. Washington, D. C.: American Institute of Biological Sciences, 1965.

(Received June 20, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 425–432

GOAL EVENTS AS DISCRIMINATIVE STIMULI OVER EXTENDED INTERTRIAL INTERVALS[1]

MARTIN PSCHIRRER[2]

Rockford College

Discrimination learning was investigated using differential reinforcers as discriminative stimuli in both a runway and a T-maze situation. In the initial experiment, *S*s (Wistar strain, male albino rats) were administered a sequence of three runway trial outcomes repeated indefinitely: either milk, pellets, and nonreinforcement (called MPN) or pellets, milk, and nonreinforcement (called PMN). A 15-min. ITI separated each of the nine daily trials. The results supported the prediction that *S*s would learn to run more slowly on N trials than on either M or P trials. In the second experiment, *S*s were trained to run to the left side of a T maze after a milk-reinforced trial and to the right side after a pellet-reinforced trial or vice versa. An ITI of 3 min. was employed. The results indicated that each of the six *S*s gradually learned to run to the correct side more frequently than the chance level of 50%. The results of these two experiments were interpreted as being consistent with E. J. Capaldi's sequential hypothesis.

The present investigation was concerned with the ability of *S*s (Wistar strain, male albino rats) to form an association between two temporally separated events in both pattern learning and learned alternation situations. In the patterning situation (Capaldi, 1967), *S* must use the reinforcement outcome of a preceding runway trial as a discriminative stimulus for the following trial. Thus, *S*s are administered reinforcement in a single alternating (SA) sequence of trials in which reinforced (R) and nonreinforced (N) trials are alternated. If *S* is to learn the discrimination, R trials must function as S− since they predict nonreinforcement on the following trial, and N trials must function as S+ since they predict reinforcement. After receiving either an R or N trial, *S* is returned to a neutral waiting cage for an intertrial interval (ITI) of up to 24 hr. After the ITI, *S* is allowed to traverse the runway once more and again receives either reinforcement or nonreinforcement, depending on what happened on the prior trial of the sequence.

The results of a typical pattern-running study (Bloom & Capaldi, 1961: Capaldi & Spivey, 1964) indicate that *S*s learn to run faster on R trials (i.e., after nonreinforcement) than on N trials (i.e., after reinforcement). There are the following evidences that pattern running is due to discrimination learning over a delay rather than any alternative mechanism, such as an innate tendency to run slower after reinforcement: (*a*) If the SA schedule of reinforcement is changed to a schedule in which either N or R trials no longer are valid discriminative stimuli, the *S*'s behavior changes appropriately. For instance, if the sequence is R–N–N repeated indefinitely, *S*s run slowly after the R trial but fast after each of the N trials because the R trial remains S− for the following trial (Capaldi & Senko, 1962). However, if the sequence is R–R–N–N repeated indefinitely, there is no patterning because neither an R nor an N trial has a consistent role as S− (Bloom & Capaldi, 1961). (*b*) The patterning effect develops gradually over trials in a manner reminiscent of the development of a discrimination.

---

[1] This article is part of a dissertation submitted to the Northern Illinois University graduate school in partial fulfillment of the requirements for the PhD degree. The writer wishes to express his appreciation to Sam Revusky, under whose direction this research was conducted, for his valuable comments and advice. The experiment reported in this article was supported partly by United States Public Health Service Grant MH16643. The preparation of the final manuscript was supported by the Mary Ashby Cheek Fund.

[2] Requests for reprints should be sent to Martin E. Pschirrer, Department of Psychology, Rockford College, Rockford, Illinois 61101.

Additional evidence that *S*s can use events of a preceding trial as discriminative stimuli for the following trial is supplied by the work of Petrinovich and Bolles (1957) on learned alternation by *S*s in a T maze. The first of three daily trials was reinforced regardless of the direction in which *S* ran; Trials 2 and 3 were reinforced only if *S* alternated its responses from the preceding trial. Thus, for instance, a run to the right was S+ for the running to the left on the following trial and S− for the running to the right on the following trial. All *S*s were started with a 15-min. ITI which was extended gradually until none of the original eight *S*s were alternating reliably. Some *S*s were able to alternate with ITIs of up to $5\frac{1}{2}$ hr.

The present series of experiments was designed to supply additional evidence that the patterning effect reported by Capaldi (1967) and the alternation effect reported by Petrinovich and Bolles (1957) are instances of discrimination learning in which the discriminative stimulus occurs at the end of the preceding trial.

EXPERIMENT I

It was hypothesized in the introduction that more complicated patterns of reinforcement, such as the double alternation, do not lead to pattern running in *S*s because neither an R trial nor an N trial has a consistent role as S−. The following experiment was designed to confirm this hypothesis by showing that *S*s could learn a complex R_1–R_2–N sequence repeated throughout the experiment if the first reinforcer of the sequence were made discriminably different from the second reinforcer. For instance, the first reinforcer (R_1) might be milk and the second reinforcer (R_2) might be chow pellets, and *S* would have to learn that nonreinforcement follows a reinforcement of chow pellets and not a reinforcement of milk. Thus, the pellets would be S− for the following trial and the milk would be S+. This is referred to here as the triple discrimination problem (TDP).

In short, it was predicted that in the sequence of milk–pellets–nonreinforcement, *S* would run slowly after a pellet trial (S−) but would run rapidly after nonreinforced and milk trials (S+). In the counterbalanced sequence of pellets–milk–nonreinforcement, it was predicted that *S* would run slowly after a milk trial (S−) but would run rapidly after nonreinforced and pellet trials (S+). This would show that particular properties of the reinforcement could become discriminative stimuli and not only the presence of reinforcement versus its absence.

In Exp. IA, *S*s were trained from the beginning on the TDP while in Exp. IB, *S*s were faded into the TDP from an SA problem. Extrapolating from findings summarized by Terrace (1963), it might be expected that this fading procedure would expedite the learning of the TDP.

Method

Subjects.—The *S*s, purchased from Windsor Biological Gardens, were 12 male albino rats of the Wistar strain about 75 to 90 days old at the beginning of the experiment. The *S*s were housed individually in the experimental room under both artificially controlled temperature and continuous artificial illumination.

Apparatus.—The apparatus, a Hunter commercial runway, was an enclosed straight runway that was uniformly 12.7 cm. high, 8.3 cm. wide, and 182.9 cm. long. The adjoining start box was 29.2 cm. long, 8.3 cm. wide, and 12.7 cm. high. An automatic guillotine door separated the start box from the actual runway. Both the start box and the runway had lucite tops and sides and aluminum rung floors. The goal box (33.0 cm. long, 15.2 cm. wide, and 12.7 cm. high) was constructed of $\frac{3}{4}$-in. plywood with a hinged, hardware cloth top and an aluminum rung floor. A manual guillotine door separated the runway from the goal box. The rear of the goal box was divided into two small adjacent compartments, each of which had its own guillotine door separating it from the rest of the goal box. Each of these individual compartments measured 6.3 cm. long, 7.5 cm. wide, and 12.7 cm. high. The runway and the start box were painted a medium gray while the goal box was painted a flat black.

A photoelectric system was used to record the time taken by each *S* to traverse successive portions of the runway. The three latency measures (i.e., start, run, and goal times) were recorded for each trial on three millisecond stop clocks. The first clock started when the start box door opened and which stopped when *S* broke the first photobeam which was located 25.4 cm. down the runway. Breaking of the first photobeam also started the second stop clock, which in turn stopped when *S* broke the second beam located 114.3 cm. from the first photobeam. The third stop clock started when *S* broke the second beam and stopped when *S* broke the third beam. The third photobeam, located 40.6 cm. from the second beam and only 2.5 cm. from the goal box, started a timer which measured the interval that *S* was confined in the goal box. A second timer was used to measure the 15-min. ITI.

Carnation evaporated milk was always in the right compartment of the goal box and Noyes pellets were always in the left compartment.

Preliminary training.—On each of the first 5 days, each *S* was handled individually for 5 min. a day and, in addition, was reduced to 80% ad-lib body weight. Water was provided ad lib in the home cage during the entire experiment. On Day 5, all rats were allowed to freely explore the runway for 5 min. each. The guillotine doors at each end of the runway were open during this period. In addition, there was no reinforcement in the goal box during this adaptation period. After being acclimated to the laboratory, each rat was trained to traverse the runway on Day 6. In this training, which took place prior to formal pattern training, each *S* was given six trials reinforced with 10% sucrose solution. One of these trials was a direct placement in the goal box and the other five trials were begun from the start box. At this point, six *Ss* were placed randomly in Exp. IA while the remaining six were placed in Exp. IB.

Pattern training: Experiment IA.—After the preliminary training was completed, the formal training procedure began. Prior to starting a trial, an *S* was placed in the start box with the guillotine door in a closed position. As soon as *S* was oriented toward the gate, it was opened electrically which in turn activated the first clock. The appropriate clocks were started and stopped as *S* traversed the runway and broke the photobeams. When *S* entered the goal box, the guillotine door dividing the runway from the goal box was closed. The *S* was manually forced to enter the goal box if it failed to traverse the runway in less than 60 sec.

For Exp. IA, six *Ss* were introduced to the TDP on the day following preliminary training. Three of these *Ss* were given a pellet trial prior to nonreinforcement (i.e., M–P–N sequence) while the other three were given a milk trial prior to nonreinforcement (i.e., P–M–N sequence). The sequence was repeated three times for a total of nine daily trials. On reinforced trials, *Ss* received 30-sec. access to either Carnation evaporated milk or Noyes pellets; on nonreinforced trials *Ss* were confined for 60 sec. in the goal box without access to food. The *Ss* were run as a squad allowing 2 min. per *S* with a 3-min. break coming after the six *Ss* in a squad had completed a trial, so that the ITI was 15 min.

Pattern training: Experiment IB.—After the pretraining procedure had been completed, three *Ss* in this part of the experiment were switched to an SA sequence consisting of pellets and nonreinforcement (i.e., a P–N sequence) while the other three *Ss* were placed on a milk and nonreinforcement sequence (i.e., a M–N sequence). All *Ss* received eight trials per day with a 15-min. ITI. In other details, Exp. IB was similar to Exp. IA.

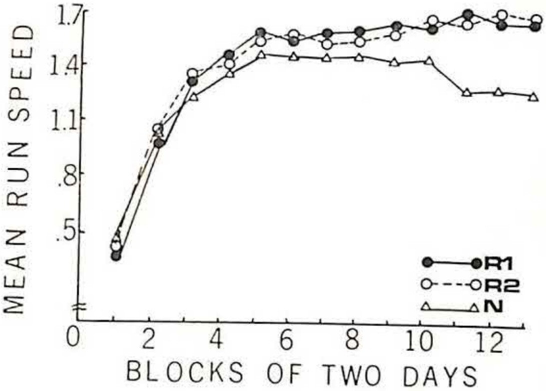


FIG. 1. Mean run speed on R_1, R_2, and N trials in a TDP sequence over blocks of 2 days.

After *Ss* had reached criterion in the above SA sequence, they were switched to the TDP. Criterion was said to have been reached for any *S* when, on 1 day, its mean running time for N trials was .5 sec. higher than its mean running time for R trials and when three of the four N trials were slower than the slowest R trial. For the P–N squad, a milk trial was added so that the sequence was milk-pellets-nonreinforcement. For the M–N squad, a pellet trial was added so that the resulting sequence was pellets–milk–nonreinforcement. Note that the same substance always preceded nonreinforcement for any given *S*. At this point in training, the procedure for *Ss* in Exp. IB was the same as that used for *Ss* in IA. The *Ss* in both experiments were tested until it was apparent that pattern running was obtained.

In order to minimize the roles of olfactory cues in Exp. I, the following three precautions were taken: First, the runway was washed thoroughly after each *S* had completed a trial. Second, the running order of *Ss* within a squad was randomized daily in an effort to eliminate the possibility that one *S* may learn the discrimination and the rest of the *Ss* merely follow the first *S*'s odor trail. Finally, *Ss* were subdivided so that two *Ss* in each experiment received nonreinforcement on either Trials 1, 4, and 7; 2, 5, and 8; or 3, 6, and 9.

Results and Discussion

Experiment IA.—Capaldi's (1967) patterning effect successfully was extended to the TDP situation. Figure 1 shows how, in the course of training, speeds on N trials became slower than speeds on R trials. (Although not presented in this article, the start and goal speeds were

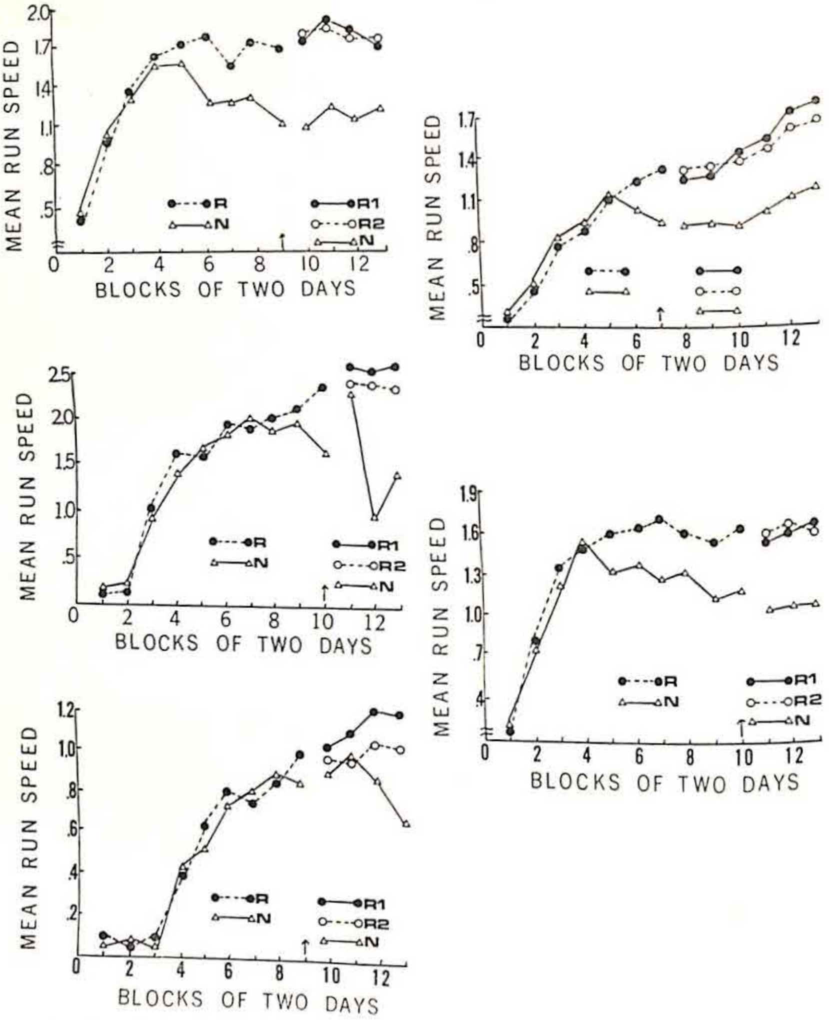


FIG. 2. Run speeds for the individual *Ss* in Exp. IB on R_1, R_2, and N trials over blocks of 2 days in the SA and TDP sequences.

found to be very similar to the run speeds.) For purposes of statistical inference, start, run, and goal speeds were averaged during Blocks 10–13; each of the six *Ss* ran faster on R trials than on N trials (binomial test, $p < .05$). To assess the consistency of the results, there were separate tests of significance on each block of trials. There was a significant difference at the .05 level for the R_1, R_2, and N run speeds during each of Blocks 8–13 (*F* tests). There was no reliable difference between the R_1 and R_2 speeds; not a single *t* test on any block yielded a significant result for any of the three measures of speed.

Figure 1 suggests that *Ss* can use the type of food reinforcement as a discriminative stimulus. Since the N trial preceded the R_1 trial and the R_1 trial preceded the R_2 trial, both N and R_1 had the roles of S+. Only R_2, which preceded the N trial, took on the role of S−.

Figure 1 also makes a subtle point. All food reinforcement is considered to have a great deal in common, but milk and pellets seem to be very different in their roles as discriminative stimuli. Consider the TDP sequence: milk (R_1), pellets (R_2), and nonreinforcement (N). Obviously, the pellets act as S− since they predict nonreinforcement on the following trial. If there were substantial generalization between the milk and the pellets, Ss would tend to run slower on R_2 trials than on R_1 trials. It is evident from Fig. 1 that nothing like this happened; thus milk and pellets were as discriminably different from each other as either is from nonreinforcement.

Experiment IB.—Since Ss were shifted individually from the SA to the TDP, the group data are very confusing and, therefore, are not shown here. Figure 2 shows the individual S data for the run section. The first part of each individual graph represents the performance on the SA pattern and the second part represents the performance of Ss on the TDP sequence. Since each of the five Ss (one died during training) learned the TDP, and in view of the results of Exp. IA, inferential statistics would add no useful information here.

Figure 2 indicates that each S began patterning at some point in the SA sequence. The individual data indicate that as soon as the fading Ss were shifted from the SA to the TDP sequence, patterning was evident *immediately* for all five Ss. This finding supports the earlier conclusion that milk and pellets are very different discriminative stimuli. If this were not the case, one might expect that after the transfer from SA to TDP, R_2 speeds initially would be about as slow as N speeds because the R_2 runs are preceded by R_1 reinforcements.

EXPERIMENT II

The purpose of Exp. II was to train Ss to run to the left side of a T maze after a milk-reinforced trial and to the right side after a pellet-reinforced trial or vice versa. In contrast to Exp. I, in which the sequence of the different types of trials was fixed, a Gellerman (1933) series of trials was used to determine the exact sequence of reinforcement on each day. If Ss were to learn successfully in Exp. II, the following new information would be obtained.

1. The transition from SA performance to TDP performance in Exp. IB was so rapid as to nearly exclude the possibility that Ss were learning to count two successive reinforcements in Exp. I without using the information supplied by the two different types of reinforcement. However, the random sequence of Exp. II insures that the problem cannot possibly be solved by counting, thus insuring that performance must depend on use of the type of reinforcement as a discriminative stimulus.

2. Experiment II requires a "go left–go right" discrimination instead of the "go–no go" discrimination of Exp. I. It is easier to postulate artifacts to account for a "go–no go" discrimination than for a "go left–go right" discrimination. Thus, if Exp. II were to be successful, it would be still more certain that the patterning effect of Exp. I is to be attributed to discrimination learning.

Method

Subjects.—The Ss were six male albino rats housed in individual Plexiglas cages in the experimental room. The same temperature, lighting, and deprivation conditions employed in earlier experiments were used in this experiment.

Apparatus.—The apparatus was a simple T maze which included a 45.7-cm. stem and two 50.8-cm. arms with goal boxes at the end of each. The goal boxes were identical to the one described in Exp. I. Except for the black goal boxes, the entire apparatus, a Hunter commercial Plexiglas T maze, was painted a medium gray. All sections of the T maze were uniformly 12.7 cm. high and 8.3 cm. wide. Three automated guillotine doors were placed at critical places in the T maze: One guillotine door separated the start box from the stem of the T maze, and one door was placed in each arm of the T maze at the choice point. The doors at the choice point were used *only* in forced trial pretraining. Both the start box and the runway had lucite tops and sides and aluminum rung floors. Manual guillotine doors when lowered separated the goal boxes from the arms of the T maze.

As in Exp. I, both compartments in each goal box were employed with Carnation evaporated milk always being in the right compartment and Noyes pellets always being in the left compartment.

Preliminary training.—On each of the first 5 days, each S was handled individually for 5 min. On Day 5, all Ss were allowed to explore the maze for 5 min. each. During this adaptation period, all five guillotine doors were open and, in addition, there was no reinforcement in the goal boxes.

Forced-trial T-maze training.—T-maze training began on Day 6 of the experiment. The rats were divided into 2 squads of 3 Ss each with the further restriction that Squad 1 (milk–right, pellets–left discrimination) was tested first each day and Squad

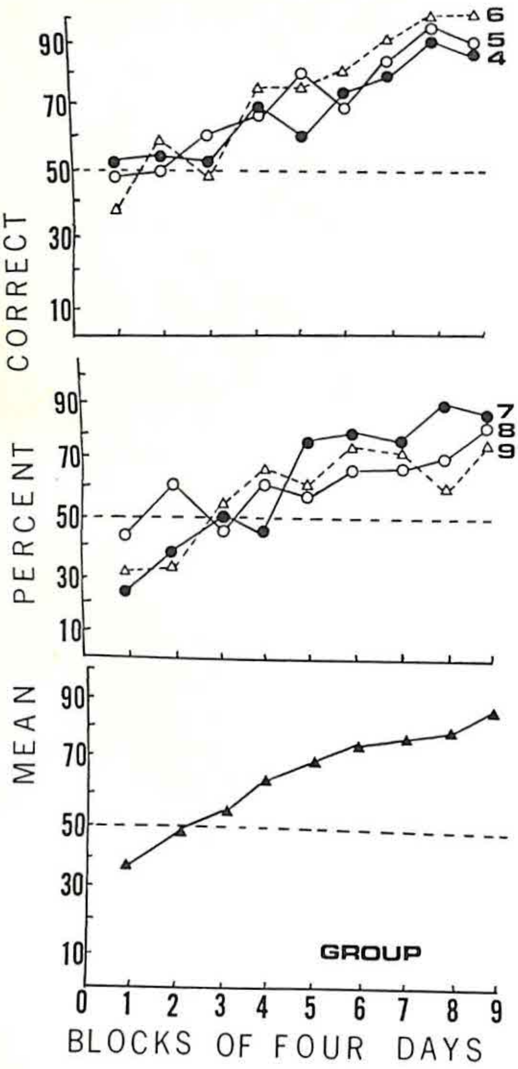


FIG. 3. Mean percent correct over blocks of 4 days for S_4–S_9 in the T-maze situation. (Squad 1 consisted of S_4–S_6, while Squad 2 consisted of S_7–S_9.)

2 (milk-left, pellets-right discrimination) was tested second. The rats within a squad were assigned randomly a daily running position which was changed from day to day.

On Days 6–10 *S*s were given a total of 40 forced trials with 20 being to the left and 20 being to the right. The sequence of trials was determined by the Gellerman series and was different for each *S* and in each day. During the forced-trial procedure, each *S* was subjected to five daily trials on Days 6 and 7 and 10 trials a day thereafter for the remaining 3 days. For example, suppose that a hypothetical *S* has to learn a discrimination in which the type of reinforcer is the relevant cue. That is, milk consumed on Trial T*n* indicates that *S* should make a right turn on Trial T*n* + 1 while pellets consumed on Trial T*n* indicates that *S* should turn *left* on Trial T*n* + 1. Thus, if *S* received pellets on Trial 1, it would be forced to the left on Trial 2; and if it received milk on Trial 1, it would be forced to the right on Trial 2. This procedure was repeated throughout forced trial training.

Free-choice T-*maze training.*—Beginning on Day 11, each *S* was given 20 daily free-choice trials for the remainder of the experiment which consisted of 720 training trials (i.e., 36 days of training). In this part of the procedure, *S* had to utilize the type of reinforcer as a discriminative stimulus if the problem was to be learned. Consider an arbitrary trial called T*n*. The correct response was determined by the type of reinforcement on the preceding trial (T*n*−1). If the rat was correct, it received immediate reinforcement in the goal box consisting of 15-sec. access to either milk or pellets as determined by the Gellerman series.

If *S* was incorrect on T*n*, a correction procedure was used since it was necessary to dispense reinforcement on each trial so that *S* would have a discriminative stimulus for the following trial. However, to prevent perseveration of strong position habits, the correction procedure included the following two methodological features: (*a*) *Delay of reinforcement after correction:* If the rat was incorrect, reinforcement not immediate after it corrected itself so that an incorrect response would not receive immediate secondary reinforcement. The *S* was confined in the *correct* goal box for 20 sec. before being given a 15-sec. reinforcement in that goal box. (*b*) *Repetition of same contingency if incorrect:* After an incorrect trial, the same type of reinforcement which had been administered on the incorrect preceding trial was administered again. That is, the rat was not advanced in the sequence of milk and pellet reinforcers until it was correct at the beginning of a trial. Thus if a rat had a position habit and reached a point in the milk–pellets sequence which required a response contrary to the position habit, it would not receive immediate reinforcement until it ran contrary to its position habit.

Each *S* in a squad was given Trial 1 before Trial 2 was begun. After the three *S*s in Squad 1 had finished their 20 daily trials, Squad 2 was tested in the same manner as Squad 1. The ITI was between 3 and 3½ min. for all *S*s. The maze was washed with a damp sponge after each squad trial and between squads.

Data analysis.—In calculating the proportion of trials in which an *S* was correct, all trials which followed incorrect trials were omitted from the data analysis. This procedure was used to insure that the data shown would measure the development of control of *S*'s behavior by the type of reinforcement received on the preceding trial. If trials following incorrect trials were included in the data analysis and if *S* learned to alternate its response following a trial on which it was forced to correct itself, performance could easily be shown as better than chance without any learning of a discrimination based on the type of reinforcement which had been received.

Results and Discussion

Figure 3 shows the proportion of runs to the correct side over blocks of 4 days for each *S* and for the group. Figure 3 indicates that each of the 6 *S*s gradually learned to run to the correct side more frequently than the chance level of 50% (binomial test, $p < .05$). The tendency of *S*s to be below chance (50%) during the first block of trials may be an artifact of the Gellerman series which has a slight tendency to reinforce alternation, while *S*s in the present situation had a tendency toward position habit perseveration.

Figure 3 shows that *S*s could use milk or pellets as cues for a directional response over a delay just as *S*s of Petrinovich and Bolles (1957) could use their own behavior as cues over a delay.

GENERAL DISCUSSION

While the present theoretical orientation is quite different from that of Capaldi (1967), the preceding results support his sequential hypothesis that it is the occurrence of R and N trials in a particular sequence which is responsible for pattern running (e.g., Bloom & Capaldi, 1961; Capaldi & Stanley, 1963).

The results of Exp. I suggest that the failure of Bloom and Capaldi (1961) to obtain DA patterning was due to the fact that no trial outcomes in this sequence consistently took on the role of $S-$ or $S+$. The *S*s tended to solve their problem by running very fast on all trials of the sequence. A study by Hart (reported by Capaldi, 1967, p. 129) further supports the hypothesis that DA patterning fails to occur because there is no consistent $S-$ or $S+$. Hart administered a sequence of $R-N_1-N_2$ trials to *S*s (rats) in a straight runway. He reported that *S*s did not differentiate between N_2 and R trials but did learn to run slowly on N_1 trials—the trials following the R trials. Note that the R trial was the only trial in the three-trial sequence which always predicted the same consequence. That is, R trials consistently had the role of $S-$, while the N trials were on some occasions $S-$ (i.e., Trial N_1) and on other occasions $S+$ (i.e., Trial N_2). It might be hypothesized that if the two N trials somehow had been made different (i.e., shock on N_1 trials vs. confinement on N_2 trials), just as the two R trials were made different in Exp. I, *S*s would have run slowly on all N trials and fast on R trials.

Experiment I indicated that patterning is quite likely a type of discrimination learning since it was found that if the R trials in a TDP are made different, *S* readily learns when the last reinforced trial has occurred and when the first nonreinforced trial will occur in the sequence. Therefore, it seems reasonable to conclude that the results of Bloom and Capaldi (1961) which indicated that *S*s were not able to learn the DA problem were probably not due to the inability to learn, but more likely were probably due to the inability of the *S*s to discriminate between the R trials of the sequence; thus there was no consistent $S-$ to predict the occurrence of an N trial. Likewise, since the N trials were not different in the DA situation, *S*s had similar trouble determining when the last N trial had occurred; thus there was no consistent $S+$ to predict the occurrence of an R trial.

Furthermore, the results of the fading group in Exp. IB suggested that the TDP situation could even be simplified to a great degree if *S*s were given training in an SA situation prior to being exposed to the more complicated TDP. In this situation, it appeared as if *S*s learned that reinforcement of one kind (i.e., milk for some *S*s and pellets for others) always indicated that nonreinforcement would follow (i.e., took on the role of $S-$). Then when *S*s were shifted to the TDP, the addition of another reinforcer did not necessarily disrupt the patterning response since *S*s still only had to utilize milk as $S-$ (or vice versa). The newly added reinforcer rapidly took on the role of $S+$ since it predicted the occurrence of reinforcement on the following trial.

The results of Exp. II extended the findings of Petrinovich and Bolles (1957) since it was demonstrated that directional responses can come under the control of differential reinforcers, even when spaced trials are employed. The results showed that the reinforcement outcome of the preceding trial has still another property which is characteristic of discriminative stimuli: It can control the type of response *S* will make on the following trial instead of just the speed of the response.

REFERENCES

BLOOM, J. M., & CAPALDI, E. J. The behavior of rats in relation to complex patterns of partial reinforcement. *Journal of Comparative and Physiological Psychology*, 1961, **54**, 261–265.

CAPALDI, E. J. A sequential hypothesis of instrumental learning. In K. W. Spence & J. T. Spence

(Eds.), *The psychology of learning and motivation.* New York: Academic Press, 1967.

Capaldi, E. J., & Senko, M. L. Acquisition and transfer in partial reinforcement. *Journal of Experimental Psychology,* 1962, **63**, 155–159.

Capaldi, E. J., & Spivey, J. E. Stimulus consequences of reinforcement and nonreinforcement: Stimulus traces or memories. *Psychonomic Science,* 1964, **1**, 403–404.

Capaldi, E. J., & Stanley, L. R. Temporal properties of reinforcement aftereffects. *Journal of Experimental Psychology,* 1963, **65**, 169–175.

Gellerman, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *Journal of Genetic Psychology,* 1933, **42**, 207–208.

Petrinovich, L., & Bolles, R. C. Delayed alternation: Evidence for symbolic processes in the rat. *Journal of Comparative and Physiological Psychology,* 1957, **50**, 363–365.

Terrace, H. S. Discrimination learning with and without "errors." *Journal of the Experimental Analysis of Behavior,* 1963, **6**, 1–27.

(Received June 23, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 433–438

SUCCESSIVE NEGATIVE CONTRAST EFFECT:

INTERTRIAL INTERVAL, TYPE OF SHIFT, AND FOUR SOURCES OF GENERALIZATION DECREMENT[1]

E. J. CAPALDI[2]

Purdue University

Employing rats in a runway, the successive negative contrast effect was studied varying both intertrial interval (3–5 min. or 24 hr.) and type of reward shift (transfer vs. nontransfer shift). The negative contrast effect was considerably larger at the 3- to 5-min. interval. Transfer shift training eliminated the negative contrast effect at the 24-hr. interval, merely reducing it at the 3- to 5-min. interval. These findings were employed to evaluate the relative contribution to the negative contrast effect of four separate sources of generalization decrement.

A reduction in reward magnitude may depress performance below that of a small reward control group, a phenomenon called the successive negative contrast effect (NCE). Some hypotheses suggest that associated with reward reduction is the production of new unconditioned stimuli which weakens responding due to generalization decrement and the elicitation of competing responses. While there has been a tendency to explain the NCE in terms of some single source of generalization decrement, it seems clear that at least two sources of generalization decrement must be involved. For one, reward reduction appears to produce frustrative stimuli (e.g., Amsel, 1958). Also, it appears (Capaldi & Lynch, 1967) that different reward magnitudes (large, small, etc.) produce associated and distinctive stimuli. Thus, an animal trained under large reward would, upon being shifted to small reward, encounter the unconditioned stimuli produced not merely by frustration but by small reward as well, and would tend to perform poorly (NCE). Recent evidence (Amsel, Wong, & Traupmann, 1971; Capaldi, Berg, & Sparling, 1971) suggests, however, that as many as four unconditioned stimuli may be involved in producing the NCE. That is, it appears that any stimulus state, such as frustration, consists of two separate and distinct components, a component which diminishes as intertrial interval (ITI) increases—many would call this component a fading trace or aftereffect—and a component which is independent of ITI. The component which is unaltered by ITI might depend upon classical conditioning (Amsel et al., 1971) or long-term memory (Capaldi et al., 1971). More particularly, one sort of frustrative stimulus (S^F) appears to be at maximal intensity immediately following the frustrative trial thereupon progressively diminishing to zero as ITI increases, while the second sort of frustrative stimulus (S^f) appears to be unaffected or unaltered by ITI. Correspondingly, the stimulus complex produced by small reward has a diminishing component (S^{RS}) and an ITI-independent component (S^{rs}). Accordingly, while reward reduction at massed trials would introduce all four unconditioned stimuli, S^F, S^f, S^{RS}, and S^{rs}, reward reduction at spaced trials would introduce only S^f and S^{rs}. Clearly, then, the NCE should be larger at massed than at spaced trials. The effects of ITI on the NCE do not seem to have been directly investigated. Spear and Spitzner (1968) have noted, however, that the NCE frequently has proved to be a fragile phenomenon, both absolutely small and difficult to demonstrate statistically. This observation may be valid only in connection with the spaced trial NCE.

The investigation reported here had two major purposes. One purpose was to determine if as many as four sources of generalization decrement are involved in producing the NCE. The second purpose is unique in NCE investigation and went well beyond the first. The second purpose was to evaluate, within reasonable limits, the relative contribution to the NCE of each of the four sources of generalization decrement noted.

---

[1] This research was supported in part by National Institute of Child Health Development Grant 5R HD 04379-03 to E. J. Capaldi.

[2] Requests for reprints should be sent to E. J. Capaldi, Department of Psychological Sciences, Purdue University, Lafayette, Indiana 47907.

As indicated, the NCE occurs because one or more of the stimuli S^F, S^f, S^{RS}, and S^{rs} are not conditioned to the instrumental reaction at the time reward is reduced. Clearly, however, if one or more of these stimuli were conditioned to the instrumental reaction prior to the reward reduction phase, then the NCE might be reduced and perhaps even eliminated. Thus, it has been reported (Capaldi & Ziff, 1969; Mikulka, Lehr, & Pavlik, 1967) that a shift to a small magnitude of consistent reward following training under a large magnitude of partial reward did not produce an NCE. The NCE was eliminated presumably because under partial reward frustrative stimuli are conditioned to the instrumental reaction in the training phase, and thus in the reward reduction phase the frustrative stimuli did not occasion generalization decrement.

Capaldi and Lynch (1967), who employed consistent reward and a 24-hr. ITI, eliminated the generalization decrement associated with stimuli characteristic of small reward, specifically S^{rs}. First, Capaldi and Lynch reported that a group which received small reward following large reward (usual two-phase design) showed an NCE. For this group, called the nontransfer shift group (NTS), both S^f and S^{rs} produced generalization decrement. A second group, called the transfer shift group (TS), was trained in three phases, small reward, large reward, and small reward again. The purpose of the initial phase of small reward trials was to condition S^{rs} to the instrumental reaction. Thus, for the TS group in the reward reduction phase, S^f but not S^{rs} produced generalization decrement. The TS group did not show an NCE, suggesting that when the ITI is 24 hr., eliminating the generalization decrement associated with S^{rs} can eliminate the NCE.

Since transfer shift training at a long ITI eliminated S^{rs} as a source of generalization decrement, transfer shift training at a short ITI should eliminate both S^{RS} and S^{rs} as active sources of generalization decrement. This investigation contained a transfer shift group trained at massed trials (Group TSM) and a transfer shift group trained at spaced trials (Group TSS), as well as a nontransfer shift group trained at massed trials (NTSM) and a nontransfer shift group trained at spaced trials (NTSS). These groups constitute a 2 × 2 factorial, TS versus NTS training combined with massed versus spaced trials. The purpose of the TS training was to eliminate the generalization decrement associated with S^{rs} at spaced trials (Group TSS) and the generalization decrement associated with S^{RS} and S^{rs} at massed trials (Group TSM). When this is done it becomes possible to estimate the relative amount of generalization decrement associated with each of the four stimuli in the reward reduction phase.

The active sources of generalization decrement in the reward reduction phase were for each of the four groups as follows: Group NTSM (S^F, S^f, S^{RS}, and S^{rs}), Group TSM (S^F, S^f), Group NTSS (S^f, S^{rs}), and Group TSS (S^{rs}). Note that there are two active sources of generalization decrement in the reward reduction phase for Group TSM (S^F and S^f) and Group NTSS (S^f and S^{rs}) and that these groups share a S^f in common. If Group TSM showed a larger NCE than Group NTSS, it would necessarily follow that more generalization decrement was associated with S^F than with S^{rs}. An opposite result would support an opposite conclusion. By comparing the size of the NCE in other groups or combinations of groups, it was possible to evaluate the relative amount of generalization decrement associated with the other stimuli.

METHOD

Subjects.—The *Ss*, about 90 days old at the start of the experiment, were 60 naive male rats purchased from the Holtzman Company, Madison, Wisconsin.

Apparatus.—The apparatus was a straight alley, 208.3 cm. long, 10.2 cm. wide, and enclosed by 22.9-cm.-high sides, covered with hinged 1.3-cm. hardware cloth. It was constructed of wood and painted a medium gray. It had three basic sections: a 35.6-cm. start section, a 117.1-cm. run section, and a 40.6-cm. goal section. On all trials, *S* was placed at the rear of a 25.4-cm. start treadle suspended over a microswitch, starting a .01-sec. clock. This clock stopped and a second started when *S* broke an infrared beam 10.2 cm. from the treadle's tip (start time). The interruption of a second beam 117.1 cm. from the first stopped the second clock (run time) and started a third clock. Thirty and one-half centimeters from the second beam and 5.1 cm. from the front edge of a brass 5.1 × 10.2 × 3.8 cm. food cup, covered by a tightly fitting, automatically controlled, sliding metal lid operated by an electric motor, was a third infrared beam. Interrupting it stopped the third clock (goal time) and opened the lid covering the food cup. When *S* broke the third beam, a brass guillotine door 30.5 cm. from the alley's distal end was lowered manually, thereby confining *S* to the goal area. Individual times for all three sections were recorded, as well as "total" time, which was the sum of the elapsed times on the clocks.

Preliminary training.—The *Ss* were placed in individual cages and fed 12 gm. of Wayne Lab Blox each day, water being available freely. On Days 2–12 of pretraining, *Ss* were handled in groups of six for 10 min. each day outside the home cage. Throughout the course of the experiment, the daily ration was fed 15 min. following return to the home cage.

Experimental training.—Experimental training may be divided into three phases: In Phase 1, groups received either 15 small reward trials (2.045-gm. Noyes pellets) or handling; in Phase 2, groups received either 30 large (20 pellets) or 30 small reward trials; and in Phase 3, all groups received 20 small reward trials. Three groups were trained at a 24-hr. ITI and three corresponding groups at a 3- to 5-min. ITI.

The small reward, spaced trial group (Group SS) received one trial each day on Days 1–45 of experimental training. The transfer shift, spaced trial group (Group TSS) received 15 small reward trials (Days 1–15) followed by 30 large reward trials (Days 16–45). The nontransfer shift, spaced trial group (Group NTSS) was removed from the home cage and handled on Days 1–15, this being followed by 30 large reward trials (Days 16–45).

Similar groups were trained at a 3- to 5-min. ITI. In Phases 1 and 2, these groups received three trials each day. Group SM (small reward, massed) received small reward on Days 31–45, Group TSM received small reward on Days 31–35 followed by large reward on Days 36–45, and Group NTSM received large reward on Days 36–45. Groups SM and TSM were removed from the home cage and handled on Days 1–30 as was Group NTSM on Days 1–35.

On Days 46–65 (Phase 3), the three spaced trial groups received small reward at one trial each day. The three massed trial groups received 20 small reward trials, 10 on Day 46 and 10 on Day 47, the ITI remaining at 3- to 5-min.

Groups TS and NTS were equated on the basis of number of large reward trials (30) rather than on the basis of total number of trials (45). The reason was that the NCE increases as number of large reward trials increase (e.g., Vogel, Mikulka, & Spear, 1966). Thus, if total number of trials were equated, finding the NCE to be larger in the NTS condition as opposed to the TS condition could be ascribed either to the greater number of large reward trials in the NTS condition or the small reward trials received in the TS condition. Since the NCE does not decrease with an increase in number of trials per se, e.g., as number of large reward trials increase the NCE increases, it was considered best to provide the TS group with 15 additional small reward trials.

The *Ss* were brought into the experimental room in squads of three in holding cages. A squad always consisted of one *S* from each of the three massed trial groups or one *S* from each of the three spaced trial groups. Massed and spaced trial squads were run alternately and the running order of *Ss* within a squad was changed daily. Immediately after having eaten the reward in the goal box, *S* was returned to the holding cage. Following the last daily trial, *Ss* were returned to the home cage where 15 min. later they were fed 12 gm. of Lab Blox minus the amount eaten in the runway. There were two *Es*. Each *E* ran half of the *Ss* in each group.

RESULTS

Analyses of variance included three main effects: *E*, massed versus spaced trials (ITI) and prior training (PT), small reward versus large reward (NTS training) versus TS training.

Phase 1.—In Phase 1, groups received either small reward trials in the runway or handling. An analysis of variance over the 15 trials of Phase 1 indicated that no difference approached significance including that due to the dummy variable (Groups SS and SM vs. Groups TSS and TSM).

Phase 2.—Since the NTS group began training in Phase 2, an analysis over the early trials of Phase 2 which includes these groups is not too meaningful. Accordingly, an analysis over the first 15 trials of Phase 2 was performed employing the small reward groups and the TS groups. Only the difference due to PT was significant. $F(2, 48) = 4.12$, $p < .05$, in start; $F = 3.78$, $p < .05$, in run; $F = 4.14$, $p < .05$, in goal; $F = 4.66$, $p < .05$, in total. In Fig. 1, the speeds were collapsed across ITI (differences due to this variable did not approach significance), and the speeds in total of the major three groups on the last 15 trials of

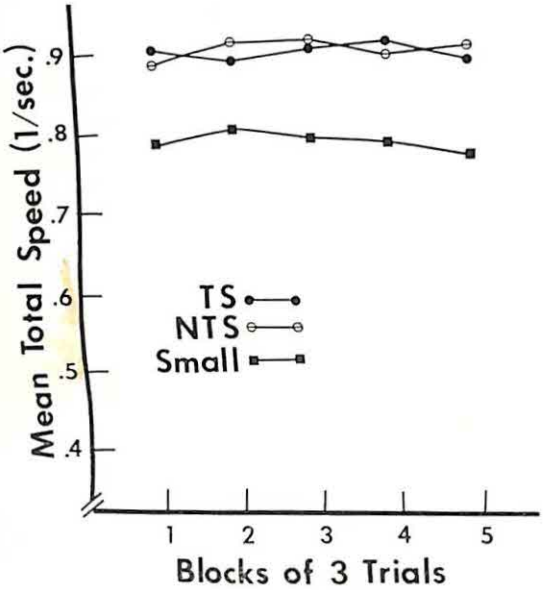


FIG. 1. Speeds in total in blocks of three trials for groups in the TS condition, the NTS condition, and the small reward condition on the last 15 trials of Phase 2.

Phase 2 in blocks of three trials are shown. It can be seen that large reward produced faster running than small reward, F (2, 48) $= 6.07$, $p < .05$, in start; $F = 7.37$, $p < .01$, in run; $F = 7.76$, $p < .01$, in goal; $F = 7.81$, $p < .01$, in total. Differences due to E did not approach significance. Of major interest in Fig. 1 is that, as was expected on the basis of previous findings (Capaldi & Lynch, 1967), all groups had reached asymptotic performance by this point in training. Thus, in terms of evaluating the NCE, providing Groups SM and SS with 45 rather than 30 trials (corresponding to the number received by Groups TSM and TSS rather than Groups NTSM and NTSS) made no practical difference.

Phase 3.—Figure 2 shows the speeds of the six groups in blocks of 4 trials over the 20 trials of Phase 3 in total. With a single exception, curves for the start, run, and goal sections resemble exactly those shown in Fig. 2; the exception being, as will become clear, that the size of the NCE increased from start to goal. It can be seen in Fig. 2 that the difference between Groups SM and SS was negligible, that the NCE was considerably larger at massed than at spaced trials, that TS training merely reduced the massed trial NCE but completely eliminated the spaced trial NCE, that the spaced trial NTS group shows only a fragile NCE, and, importantly, that Group TSM showed a larger NCE than Group NTSS.

An analysis over the 20 trials of Phase 3 indicated that while differences due to PT were not significant in start, F (2, 48) $= 2.89$, $p < .05$, they were significant in run, goal, and total, F (2, 48) $= 16.74$, $p < .01$; $F = 16.40$, $p < .01$; and $F = 21.36$, $p < .01$, respectively. The ITI × PT interaction was not significant in start, F (2, 48) $= 1.33$, $p > .05$, but was significant in run, goal, and total, F (2, 48) $= 11.82$, $p < .01$; $F = 6.67$, $p < .01$; and $F = 8.27$, $p < .01$, respectively. No other differences approached significance. Subsequent Newman-Keuls tests indicated that in run and goal but not start, Group NTSM differed significantly from all other groups. Group TSM also differed significantly from all other groups in run and goal. No other differences were significant. Thus, considering all 20 trials neither Group TSS nor Group NTSS showed an NCE.

Each trial block was analyzed separately. The analysis on the second block of trials

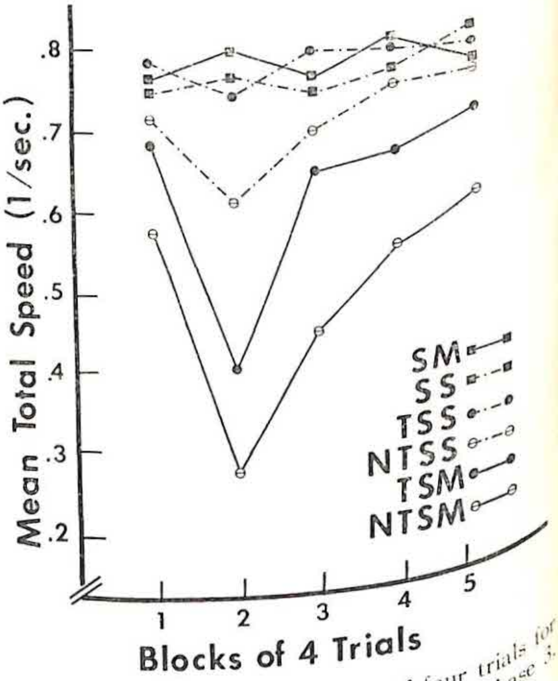


FIG. 2. Speeds in total in blocks of four trials for each of the six groups on the 20 trials of Phase 3.

gave results highly similar to those reported above with one major exception: Subsequent Newman-Keuls tests indicated that Group NTSS showed an NCE and differed from Group TSS in goal and total, not start or run. At massed trials, then, the NCE extended into the run section, included Groups TSM and NTSM, and was demonstrable employing all 20 trials of Phase 3. At spaced trials, however, the NCE occurred only in goal, did not occur in Group TSS, and was demonstrable statistically only on the second block of trials.

DISCUSSION

Since when appropriate comparisons were made, ITI did not influence speed of running either in Phase 1, 2, or 3, differences between downshift massed versus spaced trial groups may be attributed exclusively to the generalization decrement associated with each ITI, e.g., S^F, etc. It is not unusual to find, as here, that acquisition performance is relatively insensitive to acquisition ITI (see Lewis, 1956). In terms of the variables manipulated here, the major findings were that the NCE is much larger at massed than at spaced trials, and that while TS training reduced the NCE at massed trials it eliminated the NCE at spaced trials. These results suggest that the relative amount of generalization decrement associated with each of the four stimuli presented in the reward reduction phase was as follows, $S^F = S^{RS} > S^{rs} > S^f$. The specific group comparisons from which this conclusion was derived are listed below. The active sources of generalization decrement in the reward reduction phase were: for Group NTSM (S^F, S^f, S^{RS} and S^{rs}), for Group TSM (S^F, S^f), for Group NTSS (S^f, S^{rs}) and for Group TSS (S^f).

Generalization decrement associated with S^F.—The results obtained clearly suggest that more generalization decrement was associated with S^F than with S^{rs}. This becomes apparent by examining Group TSM (S^F and S^f) versus Group NTSS (S^{rs} and S^f). Both groups had two sources of generalization decrement and shared S^f in common. Since the NCE was larger in Group TSM than in Group NTSS, it necessarily follows that more generalization decrement was occasioned by S^F than by S^{rs}. Group TSS (S^f) failed to show an NCE. Thus, as in a previous investigation (Capaldi & Lynch, 1967), S^f acting alone failed to produce an NCE. Given the absence of an NCE in Group TSS (S^f) and the relatively large NCE in Group TSM (S^f and S^F), the data strongly suggest that S^F constituted a much more powerful source of generalization decrement than S^f. The NCE was larger at massed than at spaced trials, then, in part because S^F constituted a relatively powerful source of generalization decrement. Put differently, the NCE was larger at massed than at spaced trials not merely because there is at least one additional source of generalization decrement at massed trials (S^F), but also because this source of generalization decrement was more potent than that associated with either of the spaced trial sources (S^f or S^{rs}).

Generalization decrement associated with S^{rs}.—While more generalization decrement was associated with S^F than with S^{rs}, the data suggest that some generalization decrement was associated with S^{rs}. Thus, at spaced trials when the generalization decrement associated with S^{rs} was operative (Group NTSS) a small NCE occurred in the goal section, whereas when this source of generalization decrement was not operative (Group TSS) an NCE did not occur.

Generalization decrement associated with S^f.—The absence of an NCE in Group TSS suggests that either quite weak generalization decrement was associated with S^f or none at all. For reasons given elsewhere (Capaldi et al., 1971), the former alternative is preferred here. There are reasons for assuming that while S^f contributes to the NCE, under the particular experimental conditions employed in this investigation, *that* contribution went undetected. By way of example, if in this investigation small reward had been one pellet rather than two, S^f may have been more intense and an NCE might have occurred at spaced trials despite TS training.

Generalization decrement associated with S^{RS}.—In Phase 3, the speed difference between Groups NTSM and TSM (these groups differed on the basis of S^{RS} and S^{rs}) was much larger than the speed difference between Groups NTSS and TSS (these groups differed only on the basis of S^{rs}). Accordingly, the data suggest that S^{RS} substantially contributed to the NCE. It would seem that since the difference between Groups NTSM and TSM was so large in relation to the difference between Groups NTSS and TSS that S^{RS} must have constituted not merely an additional source of generalization decrement, but a more powerful source than S^{rs}. Accordingly, it would appear that in this investigation more generalization decrement was associated with either of the diminishing stimuli (S^F and S^{RS}) than with either of the ITI-independent stimuli

(S^{f} and S^{rs}). Indeed, assuming that S^{f} contributed little to the NCE (no NCE in Group TSS) it can be seen that, since the size of the difference between Groups NTSM and TSM (due primarily to S^{RS}) was about as large as that between Groups TSS and TSM (due to S^{F}), S^{RS} contributed about as much generalization decrement as S^{F}.

Are the relative amounts of generalization decrement which were found to be associated with S^{F}, S^{RS}, S^{f}, and S^{rs} peculiar to the specific experimental conditions employed here or will these relationships prove to be similar under other experimental conditions? In certain cases it seems possible that relationships somewhat but not totally unlike those reported here might be found. It might be assumed, for example, that in the reward reduction phase, frustration is an increasing function of deprivation. If this were so, then as deprivation increased more generalization decrement might tend to be associated with S^{F} than with S^{RS} and with S^{f} than with S^{rs}. In many cases, however, it would seem that variables which increased or decreased the amount of generalization decrement associated with one of the four stimuli would have the same directional effect on the remaining three stimuli. For example, the smaller the reward in the reward reduction phase, the greater might be S^{F} and S^{f} (e.g., Amsel, 1958) but also the greater would be the generalization decrement associated with S^{RS} and S^{rs} (e.g., Capaldi & Lynch, 1967). Note, however, if the ITI were reduced below the 3 to 5 min. employed here the amount of generalization decrement associated with S^{F} and S^{RS} would presumably be increased. This operation, of course, would have no effect on the generalization decrement associated with S^{f} or S^{rs} at the 24-hr. ITI. Thus it seems likely that the finding obtained here—that the amount of generalization decrement associated with S^{f} and S^{rs} is quite minor in relation to the amount of generalization decrement associated with S^{F} and S^{RS}—is general. The major sources of generalization decrement and thus of the NCE then would appear to be associated with the exclusively massed trial or diminishing stimuli (S^{F} and S^{RS}) and not the ITI-independent stimuli (S^{f} and S^{rs}). This is indeed an ironic development in as much as the role of S^{F} and S^{RS} in producing the NCE has not been previously recognized.

Bower (1961) suggested a viewpoint, since accepted by many, that phenomena such as the NCE indicate that the effects of a particular reward magnitude are not absolute but contextual. Capaldi and Lynch (1967) have suggested "Contextual hypotheses share in common the proposition that several reward magnitudes in combination generate stimuli not generated by the separate reward magnitudes [p. 226]." Since small reward in the absence of large reward would not produce either S^{F} or S^{f}, there are clearly contextual stimuli. On the other hand, S^{RS} and S^{rs} are absolute stimuli. Just as Bower suggested that the NCE is an exclusively contextual phenomena, Capaldi and Lynch suggested that the NCE is an exclusively absolute phenomena. Both views seem too extreme. That is, the hypothesis outlined here suggests that it may not be a matter of contextual versus absolute hypothesis but rather the respective roles of contextual factors (S^{F} and S^{f}) and absolute factors (S^{RS} and S^{rs}) in regulating phenomena such as the NCE.

REFERENCES

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

Amsel, A., Wong, P. R. P., & Traupmann, K. L. Short-term and long-term factors in extinction and durable persistence. *Journal of Experimental Psychology*, 1971, **90**, 90–95.

Bower, G. H. A contrast effect in differential conditioning. *Journal of Experimental Psychology*, 1961, **62**, 196–199.

Capaldi, E. J., Berg, R. F., & Sparling, D. L. Trial spacing and emotionality in the rat. *Journal of Comparative and Physiological Psychology*, 1971, **76**, 290–299.

Capaldi, E. J., & Lynch, D. Repeated shifts in reward magnitude: Evidence in favor of an associational and absolute (noncontextual) interpretation. *Journal of Experimental Psychology*, 1967, **75**, 226–235.

Capaldi, E. J., & Ziff, D. R. Schedule of partial reward and the negative contrast effect. *Journal of Comparative and Physiological Psychology*, 1969, **68**, 593–596.

Lewis, D. J. Acquisition, extinction, and spontaneous recovery as a function of percentage of reinforcement and intertrial intervals. *Journal of Experimental Psychology*, 1956, **51**, 45–53.

Mikulka, P. J., Lehr, R., & Pavlik, W. B. Effect of reinforcement schedules on reward shifts. *Journal of Experimental Psychology*, 1967, **74**, 57–62.

Spear, N. W., & Spitzner, J. H. Residual effects of reinforcer magnitude. *Journal of Experimental Psychology*, 1968, **77**, 135–150.

Vogel, J. R., Mikulka, P. J., & Spear, N. E. Effect of interpolated extinction and level of training on the "depression effect." *Journal of Experimental Psychology*, 1966, **72**, 51–60.

(Received June 23, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 439-448

VISUAL AND VERBAL CODING OF LATERALLY PRESENTED PICTURES[1]

ROBERTA L. KLATZKY[2]

Stanford University

The *S*s were required to indicate whether or not two pictures represented the same common object. On each trial, the first picture was presented in the right or left visual field, and after an interstimulus interval (ISI) of .1, .5, 1.0, or 4.0 sec., the second picture was presented in the right or left visual field. The two pictures were physically identical (an identity match), representative of the same object but not identical (a name match), or representative of different objects (a negative response). Reaction time (RT) was recorded for each response. The functions relating mean RT to ISI and functions relating the difference between name- and identity-match RTs to ISI conformed to those obtained when letters are used in a similar task (by M. I. Posner, in 1970). No difference was found between the functions corresponding to presentation of both pictures in the left visual field and presentation of both in the right field. However, a significant difference was obtained for each response when the two conditions—both pictures presented in the same visual field and each picture presented in a different field—were compared. This result was discussed in the context of a cerebral hemispheric capacity-sharing hypothesis.

Several experiments in which *S*s were required to make comparisons between two stimuli have been described by Posner and Mitchell (1967). In one variant of such tasks, *S*s were required to indicate whether or not two letters had the same name, and reaction time (RT) was recorded. There are two distinct types of positive responses that may occur in this task: those in which the two letters are physically identical (an identity match) and those in which the two letters are not identical but have the same name, e.g., a and A (a name match). Posner and Mitchell have hypothesized that these two positive matches are based on different underlying processes, namely, that identity and name matches are based on comparisons of spatial and verbal stimulus representations, respectively. A negative response is also supposed to be based on a comparison of names and is therefore quite similar to the name-matching process.

The hypothesis of Posner and Mitchell (1967) just described has received extensive empirical support (for a review, see Posner, 1970). Posner and Mitchell found that responses to identity matches were approximately 70 to 100 msec. faster than responses to name matches, showing that *S*s can in fact process identical pairs of stimuli much faster than those which have conceptual rather than physical identity. In addition, several studies (Boies, 1969; Posner, Boies, Eichelman, & Taylor, 1969; Posner & Keele, 1967) have shown that the advantage of identity matches over name matches rapidly diminishes as the interval between the two stimuli increases. In general, this advantage decays to zero over an interval of about 2 sec. (Posner, 1970). This result appears to support the spatial/verbal matching hypothesis and to indicate that identity matches are made with much more difficulty as the spatial information necessary for matching on a physical basis decays.

The exact nature of the visual code used as the basis of identity matches is somewhat open to question. However, it seems clear that this visual code is of a higher level of processing than a sensory trace of visual information (Posner, 1970). Variables such as luminance and exposure duration, and the interpolation of a noise field between the stimuli seem to affect both name and identity matches equally. These results indicate that "the basis of [an identity] match is not the trace of the stimulus, but the activation of an abstract code which serves as an internal representation . . . of prior visual experience [Posner, 1970, p. 95]."

---

[1] This research was conducted at Stanford University in the laboratory of R. C. Atkinson, whom the author wishes to thank for his assistance with all phases of this experiment. Thanks are also due to Edward Smith, who offered constructive comments on the manuscript.

[2] Requests for reprints should be sent to Roberta L. Klatzky, who is now at the Department of Psychology, University of California, Santa Barbara, California 93106.

The hypothesis that the letter-matching task just described varies in its spatial and verbal components has implications for the cerebral hemispheres involved. Specifically, the left and right hemispheres of the brain are known to be functionally asymmetrical (e.g., Bryden, 1965; Kimura, 1966; Sperry & Levy, 1970). The left hemisphere has been described as a verbal, analytic processor of information, and the right hemisphere, in contrast, has been characterized as a spatial, holistic processor. Moreover, differences in hemispheric processing have been studied in experiments that use RT as the dependent variable (e.g., Egeth & Epstein, 1972; Filbey & Gazzaniga, 1969; Geffen, Bradshaw, & Wallace, 1971; Klatzky & Atkinson, 1971). These experiments take advantage of the connections between the retina and the brain of human beings; namely, stimuli presented in one half of the visual field are directed to the contralateral hemisphere of the brain. Thus, when the stimulus duration is short enough to preclude eye movements, it may be assumed that information presented in the left visual field is transmitted to the right hemisphere, and vice versa.

In general, two hypotheses may be used to associate differences in hemispheric processing with differential RT for presentation of information to the right and left hemispheres of the brain. When RT to visually presented information is shorter for presentation in one visual field than the other, one hypothesis stipulates that the hemisphere contralateral to the field of presentation associated with the shorter RT is responsible for processing the presented information. This hypothesis assumes that when information which is directed to one hemisphere must be processed by the other, the information is transferred between the hemispheres, producing an increment in RT. Alternatively, differential RT may be explained by assuming that both hemispheres process the information, but the hemisphere associated with the shorter RT is more specialized for the processing task (and therefore more efficient) than the other hemisphere.

If the bases of identity and name matches in the task of Posner and Mitchell (1967) are in fact spatial and verbal processes, respectively, it is quite possible that the two types of matches are performed by different hemispheres. That is, one would suspect that the spatially superior right hemisphere would perform the identity matches and the verbally superior left hemisphere would perform the name and negative matches. Data supporting such a hypothesis has been reported by Cohen (1972), who used RT to assess differences in hemispheric processing in a task based on that of Posner and Mitchell. In her experiment, two letters to be compared were simultaneously presented in either the right or left visual field. Cohen found that for right-handed *Ss*, the difference between RTs for name and identity matches was less for right-field (left-hemisphere) presentation than for left-field (right-hemisphere) presentation. This is precisely what would be predicted on the basis of known hemispheric capacities, given the hypothesis that identity and name matches are based on spatial and verbal stimulus representation, respectively. That is, Cohen's data indicate that identity matches are performed by the spatially superior right hemisphere, and name matches are performed by the verbally superior left hemisphere.

In the present experiment, *Ss* were required to compare two temporally separated stimuli, each presented in the left or right visual field. Each *S* performed a series of trials of the following nature. First, a picture representing some common object (Stimulus 1) was laterally presented for 180 msec.[3] After a delay of .1, .5, 1.0, or

---

[3] In general, stimulus durations of 150 msec. or less are used in laterality studies in order to preclude eye movements; however, studies with durations greater than 150 msec. tend to be in accord with experiments using shorter durations (White, 1969). An exposure duration of 180 msec. was used in this study because pretesting indicated some difficulty in stimulus perception at shorter exposures.

4.0 sec., a second picture (Stimulus 2) was presented for 180 msec. The presentation of the second stimulus was also lateral, and the stimulus could appear either in the same position as the first stimulus or on the opposite side of the viewing area. The *S* was required to make a positive response if Stimulus 1 and Stimulus 2 represented the same object (e.g., a dog), regardless of whether or not they were identical or occupied the same position in the visual field, and a negative response if the two stimuli represented two different objects. Reaction time, defined as the interval between the onset of Stimulus 2 and *S*'s response, was recorded on each trial.

In the present experiment, as in that of Posner and Mitchell (1967), two types of positive responses could occur: identity matches, in which the two objects were identical, and name matches, in which the two objects had the same name but were not physically identical. Thus, this article provides an extension of the letter-comparison studies to the case of less familiar picture stimuli. This study also investigates effects of lateral presentation of the stimuli to be compared. These effects are of two types: First of all, a comparison may be made between RT for presentation of both stimuli in the left visual field and RT for right-field presentation of both stimuli. This comparison provides a means of investigating hemispheric asymmetries in the picture-matching task. In addition, it is possible to compare RT for the case in which the two stimuli are presented in the same visual field (i.e., both stimuli in the right field or both in the left field) with the case in which Stimulus 1 and Stimulus 2 appear in opposite fields. This comparison may be used to assess differences in comparisons between two stimuli presented to the same hemisphere and stimuli which can only be compared after interhemispheric transfer has occurred. Finally, because the interval between the two stimuli to be compared was varied in the present experiment, it is possible to investigate changes in RT and the effects of lateral presentation over time, in accordance with changes in the rate of identity matches which have previously been found (e.g., Posner et al., 1969).

METHOD

Subjects.—The *S*s were 10 female students associated with Stanford University for the summer quarter, 1971. All *S*s were right-handed, native speakers of English, and had either 20:20 vision or vision corrected to 20:20 by contact lenses. The *S*s were paid $2.00 for each experimental session.

Stimuli.—The stimuli used in this experiment were pictures (drawings) of familiar objects (e.g., a dog, snake, lion, etc.). Each picture was approximately 1¼ in. in diameter. Eighteen objects were represented in the pictures, and for each object there were 3 stimulus representations (e.g., 3 different snake pictures, 3 dog pictures, etc.), for a total of 54 pictures.

Each stimulus was mounted on a 9 × 6 in. white card so that the center of the picture was located 1⅝ in. to the right or left of the center of *S*'s viewing area and equidistant from the top and bottom (i.e., the near edge of the picture was 1 in. from the center of *S*'s visual field). For each stimulus, two cards were composed with the picture to the right of the center of the viewing area, and two cards with the picture to the left.

Apparatus.—The apparatus consisted of a three-field Iconix tachistoscope and exposure box. The stimulus presentations were controlled by *E*, who sat opposite *S*. The *S* placed her head against a head mask with her chin on a chin rest and looked into the exposure box. The viewing distance was approximately 30 in. for each field. Thus, the near edge of each stimulus was approximately 2° from the center of *S*'s visual field.

On the table in front of *S* were three telegraph keys arranged in an arc with the centers separated by a distance of 1.5 in. The *S* rested her arm on the table and depressed the keys with her forefinger. One-half of the *S*s were randomly chosen to use the key on the right for a positive response and the key on the left for a negative response, and the remaining *S*s used the reverse procedure. Each *S* was instructed to depress the center key until she was ready to respond, so that her hand position was not biased in favor of either of the response keys and, by observing *S*'s hand, *E* could insure that this instruction was followed.

Procedure.—Each *S* participated in a practice session and eight test sessions of 128 trials each during an 11-day period. Within each session, the interstimulus interval (ISI) was held constant over blocks of 8 trials and changed after each block, and no ISI was repeated until all other ISIs had been used in intervening 8-trial blocks. The sequence of ISIs was arranged according to a Latin square design over sessions and *S*s, and *S* was told which ISI would be used before each block of 8 trials. Within each session, the sequence of ISIs was held constant. Three practice trials preceded each session; the ISI used for these trials was the same as the ISI used

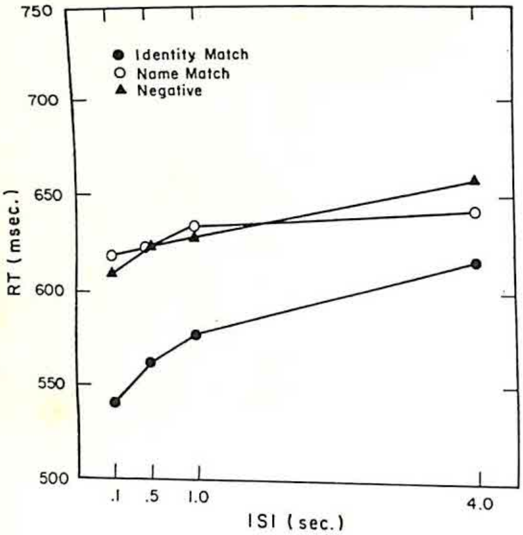


FIG. 1. Mean RT as a function of interstimulus interval (ISI) for each response condition (identity match, name match, and negative), averaged over *S*s and hemispheres of presentation and test.

during the first 8-trial block. The *S* responded with her right hand for half of each session and her left hand for the other half, and the order in which hands were used was also counterbalanced over sessions and *S*s.

The variables, namely, type of response (positive identity match, positive name match, or negative), field of presentation of Stimulus 1 (left or right), field of presentation of Stimulus 2 (left or right), and ISI (.1, .5, 1.0, or 4.0 sec.), were arranged in a factorial design for a total of 48 different trial types. For each session, within each ISI, each type of positive trial occurred equally often, and positive and negative trials occurred equally often, and the types of trials were randomly mixed. Within the constraints that each picture was used as Stimulus 1 approximately equally often and those constraints imposed by the type of trial, the specific pictures used on each trial were randomly chosen.

At the start of each session, *S* viewed and named all the pictures. Each trial consisted of the following sequence of events: (*a*) *S* pressed a start button which she held in her nonresponding hand. (*b*) A black dot on a white ground appeared in the center of *S*'s viewing area for 1 sec., and *S* fixated on the dot. (*c*) The card containing Stimulus 1 flashed on the screen for a duration of 180 msec. (*d*) The stimulus vanished; the black dot reappeared for .1, .5, 1.0, or 4.0 sec.; and *S* fixated on the dot. (*e*) The dot vanished and Stimulus 2 appeared for 180 msec. (*f*) Using her response hand, *S* pressed the key for a positive response if the two stimuli had the same name (i.e., represented the same object) or the negative key if they did not. (*g*) *E* informed *S* if she had made an error and then inserted the stimulus cards for the next trial. The interval between trials was approximately 15 sec. within an eight-trial block and 30 sec. between blocks.

RESULTS

The principal data of this experiment are mean reaction times in milliseconds. The data analysis includes correct responses only; however, the error rates for this experiment were quite low, with a range over *S*s of .6% to 3.2% and a mean of 1.7%. In addition, data for the practice session and for the three preliminary trials which preceded each test session were also excluded from the analysis. The data are averaged over right- and left-hand responses because the mean difference between hands was quite small (with the right hand faster by an average of 8 msec.), and there appeared to be no effect of responding hand on the results.

In describing the results, the following conventions are adopted: The term "presentation" refers to that phase of the trial in which Stimulus 1 is presented; "test" refers

TABLE 1

MEAN REACTION TIMES, AVERAGED OVER ALL INTERSTIMULUS INTERVALS, FOR EACH COMBINATION OF PRESENTATION AND TEST HEMISPHERES, EACH RESPONSE CONDITION, THE MEAN OVER SUBJECTS, AND EACH GROUP OF SUBJECTS

Presentation hemisphere	Test hemisphere	All *S*s			Yes-on-right *S*s			Yes-on-left *S*s		
		Identity match	Name match	Negative	Identity match	Name match	Negative	Identity match	Name match	Negative
Right	Right	578	631	628	529	572	577	626	690	678
Left	Left	574	630	645	486	552	579	662	708	711
Left	Right	574	628	626	511	563	572	638	693	680
Right	Left	572	630	624	492	552	564	651	708	684

Note.—Response conditions are identity match, name match, and negative; and *S* groups are yes-on-right and yes-on-left.

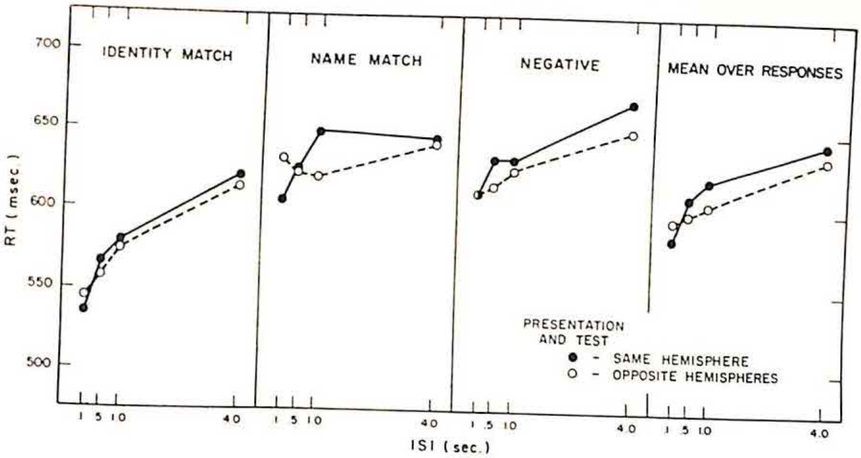


FIG. 2. Mean RT as a function of interstimulus interval (ISI) for each response condition (identity match, name match, and negative) and the mean over responses, for presentation and test in the same hemisphere (closed circles) and opposite hemispheres (open circles), averaged over *S*s.

to the Stimulus 2 phase of the trial; "field" refers to the visual field (left or right) in which a stimulus is located; and "hemisphere" refers to the hemisphere to which a stimulus is directed. (Thus, the terms field and hemisphere may be used somewhat interchangeably, with left field corresponding to right hemisphere and right field to left hemisphere.) In addition, the four conditions of presentation and test are referred to as the hemispheres to which the stimuli are directed. Specifically, the conditions are: presentation and test stimuli sent to left hemisphere (left–left), presentation and test stimuli sent to right hemisphere (right–right), presentation stimulus sent to left hemisphere and test stimulus to right hemisphere (left–right), and presentation stimulus sent to right hemisphere and test stimulus to left hemisphere (right–left).

Figure 1 presents mean RT as a function of ISI for each response condition (identity match, name match, and negative). In the figure, data are averaged over *S*s and fields of presentation and test. Thus, the data of Fig. 1 may be compared to the results which would be obtained in similar tasks in which fields of presentation and test were not varied (Posner, 1970).

In Table 1, RT is presented for each combination of presentation and test hemispheres (right–right; left–left; left–right; and right–left). Separate RTs are given for each response (identity match, name match, and negative). Results are presented for the mean over *S*s and, in addition, are broken down into two groups of *S*s: those whose positive and negative response keys were located to the right and left of the center key, respectively (yes-on-right *S*s), and those who used the reverse response key arrangement (yes-on-left *S*s). Each RT is averaged over all values of ISI.

Figure 2 presents mean RT as a function of ISI for each response condition and averaged over responses, within the conditions of presentation and test in the same hemisphere (averaged over the left–left and right–right conditions) and presentation and test in opposite hemispheres (averaged over the left–right and right–left conditions). Over all ISI values greater than .1 sec., RT for comparisons of stimuli presented and tested in opposite hemispheres is significantly less than RT when the two stimuli are presented and tested in the same hemisphere (by sign test, $p < .02$ for each response). When the ISI is .1 sec., there is no significant difference between the two conditions for any response (identity match, name match, or negative).

DISCUSSION

In general, the results shown in Fig. 1 conform to those obtained when letter stimuli are used in a similar paradigm (Posner, 1970). Reaction time for identity matches is less than RT for name matches and negative responses when the ISI is .1 sec. Although there is some increase in RT with increasing ISI for all three responses, this increase is less for name matches and negative responses than for identity matches.

A principal focus of this experiment is the investigation of hemispheric asymmetries in the matching task. According to Posner's (1970) theory, identity and name matches are based on spatial and verbal stimulus characteristics, respectively. Such matches may therefore be conducted by different hemispheres, namely, the right hemisphere for identity matches and the left for name matches as Cohen (1972) has suggested. This hypothesis may be investigated for the picture-comparison task of the present experiment by comparing the RTs given in Table 1 for the right–right and left–left conditions.

In contrast to the results of Cohen (1972), inspection of Table 1 reveals little evidence for hemispheric asymmetries in the present experiment. For the mean over *S*s of each response condition, there is no significant difference between the RTs for the right–right and left–left conditions. The reason for this result becomes apparent from further examination of the table. When *S*s are separated into yes-on-right and yes-on-left groups, a clear difference between the two groups is obtained with respect to the direction of the effect of lateral presentation. Namely, *S*s with the positive response key on the right show a strong left-hemisphere advantage, and *S*s with the positive key on the left show a right-hemisphere advantage. In fact, it is doubtful that the observed advantage for each group is truly a hemispheric effect. Instead, the results appear to be effects of compatibility; specifically, compatibility of the direction of the response with the visual field of the stimuli. That is, *S*s having the positive key on the right respond faster when the two stimuli to be compared are presented in the right visual field (a compatible response), than when the two stimuli appear in the left visual field (an incompatible response), particularly when the two stimuli are identical. For yes-on-left *S*s, this situation is reversed. These visual field advantages in turn suggest hemispheric advantages which are actually due to compatibility. In the case of the negative responses, the effects of stimulus field and response compatibility are less clear; however, the negative responses are made with the left and right response keys for yes-on-right and yes-on-left *S*s, respectively. As a result, the compatibility effect might reverse in the negative response case, and this reversal may be quite different for the two groups of *S*s.[4]

Compatibility effects like those obtained in the present experiment have been investigated in a series of experiments by Simon and his associates (e.g., Craft & Simon, 1970; Simon, 1968, 1969; Simon & Craft, 1972; Simon & Small, 1969). They found that *S*s show a strong, stereotypic tendency to respond toward the source of stimulation, even when stimulus location is irrelevant to the experimental task. When the source of stimulation and the response direction are not in correspondence, the stereotypic reaction interferes with the response, increasing RT. In the present experiment, lateral presentation of the pictures, although irrelevant to the matching task, appears to produce such a stereotypic tendency; and this tendency apparently underlies the observed compatibility effects.

The data of Table 1 indicate that the response compatibility effect is most dependent on the field of the test stimulus. In terms of the presentation and test conditions, the pattern of RTs for the right–left condition resembles the left–left pattern, and the RTs for the left–right condition resemble those of the right–right condition. This finding is not surprising, in view of the fact that the test stimulus is presented immediately prior to the response, and its field may therefore be expected to have the greatest effect on that response. Moreover, the magnitude of the compatibility effect is less when presentation and test stimuli appear in different visual fields than when the fields of the two stimuli are the same. This may be accounted for by supposing that two competing compatibility effects are operative when the stimuli occupy different fields, whereas in the same-field condition, the two compatibility effects operate in the same direction. Such an explanation is consistent with the finding of Simon, Craft, and Small (1970) that stereotypic responses toward a stimulus source can be markedly reduced by the introduction of a competing stimulus.

---

[4] An additional difference between the two groups is that of the mean RT for each group. Yes-on-left *S*s have a greater mean RT than yes-on-right *S*s, such that four out of five *S*s in the former group have mean RT above the median over all *S*s, and four out of five *S*s in the latter group have below-median RT. There is no obvious reason for this difference in mean RT, but the distribution suggests that the response key arrangement is advantageous for the yes-on-right *S*s.

The effects of stimulus field and response compatibility appear to negate any effects of hemispheric function on RT that might otherwise be observed in this experiment. Compatibility affects not only the functions relating RT to ISI for each response condition, but also the functions relating the difference between identity- and name-match RT to ISI. Averaged over ISIs, the RT difference for compatible responses (i.e., the average of the left–left condition for yes-on-right *S*s and the right–right condition for yes-on-left *S*s) is 65 msec., and this may be compared to the average difference for incompatible responses (i.e., the average of the right–right condition for yes-on-right *S*s and the left–left condition for yes-on-left *S*s) of 44 msec. Thus, the difference between RT for name and identity matches is greater for compatible than for incompatible responses. This result occurs not only for the means over ISIs, but is also found separately for each ISI. Apparently, this occurs because compatibility affects identity-match RT more than name-match RT, causing a greater difference between the two RTs for the compatible responses. Again, the relatively greater effects of compatibility appear to outweigh any hemispheric effects that might otherwise be found.

A second laterality effect examined in this experiment is concerned with the role of interhemispheric transfer in information processing. When the two stimuli to be compared are presented to different hemispheres, it is clear that information about the stimuli must be transferred between the hemispheres before comparison can take place. Such interhemispheric transfer need not take place if the two stimuli are directed to the same hemisphere. Conceivably, interhemispheric transfer of the stimuli could affect stimulus quality and or the operation of the processing system. By comparing the RTs for comparisons of stimuli presented to the same and different hemispheres, it should be possible to assess the effects of stimulus transfer on the comparison process. The relevant data, presented in Fig. 2, do not suggest that transfer of information from one hemisphere to another has a degrading effect on that information. In fact, at ISI values greater than .1 sec., RT is less for presentation and test of the two stimuli in different hemispheres than for same-hemisphere presentation and test. It is difficult to explain the significant effects of identical versus opposite hemispheres of presentation and test at ISIs greater than .1 sec. in terms of retinal fatigue, since the difference appears with ISIs as great as 4.0 sec. (However, at an ISI of .1 sec., factors such as masking, fusion, or apparent motion may have a substantial effect on response latency.) The data of Fig. 2 therefore suggest that under some conditions, presentation of two stimuli to different hemispheres facilitates comparisons of those stimuli.

The finding that presentation and test of information in opposite hemispheres may facilitate response speed is somewhat surprising, but not novel. Davis and Schmit (1971) conducted two experiments designed to explore the effect of interhemispheric transfer on information analysis. In their experiments, *S*s were required to indicate whether or not two simultaneously presented stimuli (each consisting of the digit 1 or 0) were identical, and RT was recorded. Either both stimuli were presented to the same hemisphere (the right or the left) or each stimulus was presented to a different hemisphere. They found no difference in RT for right- versus left-hemisphere presentation; however, RT was approximately 20 msec. faster for presentation of the stimuli in different hemispheres than for same-hemisphere presentation of the digit pair. These results are in agreement with the data of the present experiment. Davis and Schmit suggest that their results may reflect a refractory effect in the same-hemisphere condition. Specifically, if stimulus analysis may be done equally well by both hemispheres (as suggested by the absence of a right- vs. left-hemisphere effect), two stimuli arriving at the same hemisphere could both be analyzed in that hemisphere, producing interference between the two analyses. Such interference could cause an RT increment for the same-hemisphere condition. However, this hypothesis cannot account for the data of the present experiment, since the use of nonsimultaneous stimulus presentation means that the two pictures to be compared arrive at the hemispheres at different times. This would seem to preclude a refractory effect of the sort described by Davis and Schmit because analysis of the first stimulus could occur during the ISI, especially at an ISI of 4 sec.

Other experiments provide further evidence for a facilitative effect of stimulus presentation in different hemispheres. Dimond (1970) has found that when two visual stimuli, each directing the response of a single hand, are presented to the same hemisphere (via different eyes), response latency is much greater than when the two hands receive directions from different hemispheres (via different eyes). Also, report accuracy for word pairs appears to be greater when the two words are projected to different hemispheres than when the two words are projected to the same hemisphere (Dimond, 1971). In discussing these results, Dimond (1971, p. 185) refers to simultaneous presentation of two visual stimuli to the same hemisphere as a "block to function." A related idea is that of Gazzaniga (1970), who has proposed that one hemisphere may act as a reverberatory system while the other processes information. This leads to the idea that the hemispheres may act as a processing-sharing system, allowing two hemispheres to operate together more efficiently than one. Further evidence for sharing an information-processing load comes from an experiment by Klatzky and Atkinson (1971), in which the data suggested that while one hemisphere was retrieving and scanning information, the other was responsible for monitoring the scanning process and controlling the response. Another experiment which shows evidence for such sharing indicates that performance on a short-term memory task may be facilitated when the test stimulus is presented to the hemisphere in which initial storage was weaker (Klatzky, 1972).

The idea that the two hemispheres may act as an integrated processing system may provide an explanation for the data of Fig. 2. When two stimuli are presented to different hemispheres, the system may operate more efficiently than when the stimuli reach the same hemisphere. Even when the interval between the two stimuli is relatively long, the first stimulus must be actively retained in memory in order for comparisons to be made, and processing is facilitated by presentation of the second stimulus to the hemisphere which did not directly receive the first stimulus. Such a hypothesis can account for the fact that RT for comparisons of stimuli presented to different hemispheres is less than RT for comparisons of stimuli presented to the same hemisphere.

The finding that RT for comparisons between two stimuli is less when the stimuli are presented at different locations (i.e., to different hemispheres) than when the locations are identical (i.e., same-hemisphere presentation and test) does not necessarily violate the assumption that identity matches are based on visual stimulus representations. Since Posner (1970) assumes that matches based on visual representations are of a higher order than the stimulus trace, differences in stimulus location need not impair the comparison process. In fact, this result supports the hypothesis that identity matches are not based on the stimulus trace and thus accords quite well with data showing that luminance, exposure duration, and presence or absence of a masking field do not differentially affect name and identity matches (Posner, 1970).

It should be noted that the differences between RT for presentation and test to the same hemisphere (averaged over right–right and left–left) versus presentation and test to opposite hemispheres (averaged over right–left and left–right) cannot easily be explained by the effects of compatibility. First, the difference is found when both compatible and incompatible responses are averaged together, as in Fig. 2. In addition, evidence for the superiority of the opposite-hemisphere presentation and test can be found by examining the functions relating mean RT to ISI for compatible and incompatible responses within the opposite- and same-hemisphere conditions.

When RT is averaged over all responses and ISIs of .5, 1.0, and 4.0 sec., the mean RTs for the compatible opposite-hemisphere, incompatible opposite-hemisphere, compatible same-hemisphere, and incompatible same-hemisphere conditions are 606 msec., 618 msec., 612 msec., and 638 msec., respectively. Thus, same-hemisphere RT is greater than opposite-hemisphere RT within both the compatible and incompatible response conditions. This effect is not surprising for the incompatible condition because incompatibility produces a greater increase in RT for the same-hemisphere condition than for the opposite-hemisphere condition. However, the facilitative effect on RT of compatible responses is also greater for the same-hemisphere condition. A prediction based solely on compatibility would therefore place same-hemisphere RT less than opposite-hemisphere RT within the compatible condition, but opposite-hemisphere RT is actually less than same-hemisphere RT for compatible responses. This result supports the conclusion that the difference between the same- and opposite-hemisphere conditions does not depend on response compatibility.

The data of the present experiment may be used not only to assess hemispheric effects in the picture-matching task, but also for purposes of comparison with data from similar experiments in which letters serve as stimuli. Posner (1970) reported data for positive responses of an experiment with letter stimuli in which ISIs varied from 0 to 2 sec. and were run in blocks. Comparison of that experiment with the data of Fig. 1 indicates that the RT values for positive matches with picture stimuli are from 75 msec. to more than 100 msec. greater than the RT values for letters. Part of this difference in RT for pictures and letters is undoubtedly the effect of lateral presentation. In fact, the effect of lateral presentation can be estimated as 50 msec.[5] From this estimate, the difference between the RT values for pictures and letters appears to be partially attributable to lateral presentation. However, other variables such as luminance of the visual field and stimulus duration also differ for the two experiments, and the residual RT difference may be the result of such factors. It is also possible that the difference between RTs for letter and picture comparisons is in part a result of the nature of the stimuli. Since the pictures used in the present experiment were more complex and less familiar than letters, a relative difficulty in picture encoding may account for the greater RT for picture comparisons.

Variables such as luminance and duration may be assumed to affect name and identity matches equally (Posner, 1970), and there is little reason to suspect that the mean effect of lateral presentation (averaged over the right and left visual fields) differs for the two types of matches. In fact, the finding that changes in stimulus location do not impair identity matches (as seen in Fig. 2) supports the assumption that lateral presentation need not affect identity-match RT more than name-match RT. Thus, when the difference between RT for name and identity matches is calculated for the present experiment and similar experiments with letter stimuli (Posner, 1970), the data should be roughly comparable. For the present study, name-match RT minus identity-match RT is 78 msec., 62 msec., 56 msec., and 26 msec. for ISI values of .1 sec., .5 sec., 1.0 sec., and 4.0 sec., respectively (averaged over all presentation and test conditions and over *S*s). For experiments with letter stimuli, the difference between name- and identity-match RTs usually decays to zero by an ISI of 2 sec. (Posner, 1970). Comparison of these results indicates that the function relating the difference between name-match and identity-match RTs to ISI decreases more slowly for pictures than for letters. It has been suggested (M. Kinsbourne, personal communication, March 1972) that this function may reflect the extent to which the stimuli are verbalized. Since letters are almost invariably encountered in a reading context, it may be difficult for *S*s to make comparisons on the basis of the physical forms, especially if the ISI allows a verbal label to be generated. This difficulty is less expected in the case of picture stimuli. The physical representation of a picture may be usefully maintained over a longer interval, leading to a slower decrease in the function for pictures relative to letters.

The results of this experiment have implications for cerebral hemispheric processing and for the nature of the letter-matching task of Posner and Mitchell (1967). An important finding is that presentation of two stimuli in different visual fields facilitates the comparison of those stimuli. With respect to hemispheric function, the facilitation suggests that by sharing the information-processing load, two hemispheres may work together more efficiently than a single hemisphere. In addition, the finding that presentation of two stimuli in different locations does not impair their comparison supports the assumption that identity matches in the matching task are not based on sensory traces of the stimuli. In general, the present data from picture comparisons accord with results of similar experiments using letters (Posner, 1970), and those differences which exist may reflect differences between letters and pictures of familiar objects.

REFERENCES

Boies, S. J. Rehearsal of visual codes of single letters. Unpublished master's thesis, University of Oregon, 1969.

Bryden, M. P. Tachistoscopic recognition, handedness, and cerebral dominance. *Neuropsychologia*, 1965, **3**, 1–8.

---

[5] The effect of lateral presentation on RT may be assessed by comparing the data of two experiments in which the same pictures were used as stimuli (Klatzky & Atkinson, 1971; Klatzky, Juola, & Atkinson, 1971). In both of these experiments, *S* first memorized a set of letters. A picture was then presented, and *S* indicated whether or not the first letter of the name of that picture was contained in the memorized set. RT was recorded and plotted as a function of the number of memorized letters. In one experiment (Klatzky et al., 1971), the picture was located at the fixation point; in the other (Klatzky & Atkinson, 1971), the picture was laterally presented. Comparison of the RT functions for the two experiments reveals no difference in the slopes and a difference of about 50 msec. for the intercepts. The intercept of the RT function may be assumed to include the time required to encode the stimulus (Sternberg, 1969); thus, the 50-msec. difference between the intercepts of the two experiments provides an estimate of the effect of lateral presentation on encoding time.

Cohen, G. Hemispheric differences in a letter classification task. *Perception & Psychophysics*, 1972, **11**, 139–142.

Craft, J. L., & Simon, J. R. Processing symbolic information from a visual display: Interference from an irrelevant directional cue. *Journal of Experimental Psychology*, 1970, **83**, 415–420.

Davis, R., & Schmit, V. Timing the transfer of information between hemispheres in man. *Acta Psychologica*, 1971, **35**, 335–346.

Dimond, S. J. Hemisphere refractoriness and the control of reaction time. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 610–617.

Dimond, S. J. Hemisphere function and word registration. *Journal of Experimental Psychology*, 1971, **87**, 183–186.

Egeth, H., & Epstein, J. Differential specialization of the cerebral hemispheres for the perception of sameness and difference. *Perception & Psychophysics*, 1972, **12**, 218–220.

Filbey, R. A., & Gazzaniga, M. S. Splitting the normal brain with reaction time. *Psychonomic Science*, 1969, **17**, 335–336.

Gazzaniga, M. S. *The bisected brain*. New York: Appleton-Century-Crofts, 1970.

Geffen, G., Bradshaw, J. L., & Wallace, G. Interhemispheric effects on reaction time to verbal and nonverbal visual stimuli. *Journal of Experimental Psychology*, 1971, **87**, 415–422.

Kimura, D. Dual functional asymmetry of the brain in visual perception. *Neuropsychologia*, 1966, **4**, 275–285.

Klatzky, R. L. Cerebral hemisphere specialization and the processing of information in short-term memory. Unpublished doctoral dissertation, Stanford University, 1972.

Klatzky, R. L., & Atkinson, R. C. Specialization of the cerebral hemispheres in scanning for information in short-term memory. *Perception & Psychophysics*, 1971, **10**, 335–338.

Klatzky, R. L., Juola, J. F., & Atkinson, R. C. Test stimulus representation and experimental context effects in memory scanning. *Journal of Experimental Psychology*, 1971, **87**, 281–288.

Posner, M. I. Abstraction and the process of recognition. In J. T. Spence & G. H. Bower (Eds.), *Advances in learning and motivation*. Vol. 3. New York: Academic Press, 1970.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology*, 1969, **79** (1, Pt. 2).

Posner, M. I., & Keele, S. W. Decay of visual information from a single letter. *Science*, 1967, **158**, 137–139.

Posner, M. I., & Mitchell, R. F. Chronometric analysis of classification. *Psychological Review*, 1967, **74**, 392–409.

Simon, J. R. Effect of ear stimulated on reaction time and movement time. *Journal of Experimental Psychology*, 1968, **78**, 344–346.

Simon, J. R. Reactions toward the source of stimulation. *Journal of Experimental Psychology*, 1969, **81**, 174–176.

Simon, J. R., & Craft, J. L. Reaction time in an oddity task: Responding to the "different" element of a three-light display. *Journal of Experimental Psychology*, 1972, **92**, 405–411.

Simon, J. R., Craft, J. L., & Small, A. M., Jr. Manipulating the strength of a stereotype: Interference effects in an auditory information-processing task. *Journal of Experimental Psychology*, 1970, **86**, 63–68.

Simon, J. R., & Small, A. M., Jr. Processing auditory information: Interference from an irrelevant cue. *Journal of Applied Psychology*, 1969, **53**, 433–435.

Sperry, R. W., & Levy, J. Mental capacities of the disconnected minor hemisphere following commissurotomy. Paper presented at the meeting of the American Psychological Association, Miami, September 1970.

Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457.

White, M. J. Laterality differences in perception: A review. *Psychological Review*, 1969, **72**, 387–405.

(Received June 26, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 449–454

INFLUENCE OF RETRIEVAL CUES AND SET ORGANIZATION ON SHORT-TERM RECOGNITION MEMORY[1]

CHRISTINA A. KAMINSKY AND DONALD V. DEROSA[2]

Bowling Green State University

The *S*s were shown a series of six to-be-remembered items (positive set) composed of two subsets (letters and digits) followed by a test or probe stimulus. If the probe was a member of the series, *S*s were instructed to say "yes," and "no" if otherwise. Reaction time (RT) increased with increases in the size of the probed subsets. When positive sets were followed by an informative (I) cue, *S*s utilized a directed search strategy and confined their search to only the probed subset. When cues were noninformative ($\bar{I}$), a majority of *S*s used a nondirected search which elevated RT and decreased the slope of the function relating RT to size of the probed subset. Grouping of subsets in the positive set facilitated RT only when accompanied by informative retrieval cues.

The present investigation was concerned with an individual's ability to retrieve organized information from short-term memory (STM) in a recognition reaction time (RT) situation. Most recognition studies using RT as a dependent measure have used a single set of homogeneous to-be-remembered items (DeRosa, 1969; Sternberg, 1966). In these studies, a short series of items, typically single digits, is displayed and followed by a probe or test stimulus. When the probe appears, *S* makes one response if the probe was contained in the memorized series (positive set) and an alternative response otherwise. When RT is examined as a function of the size of the positive set, it is seen to increase at a rate of approximately 30–40 msec. per unit increase in size of the positive set. This result is taken by Sternberg (1966) as evidence of a serial search through the positive set. Furthermore, the search is presumed to be exhaustive since the rate of search is identical for positive probes and negative probes. If the search terminated when a match was made, positive probes should, on the average, result in fewer comparisons than negative probes since negative probes would always require an exhaustive search of memory.

In the present investigation, the size of the positive set was always held constant at six elements; however, the composition of the positive set was manipulated, in that positive set elements were drawn from two distinct categories—letters and digits. The positive set then contained two *subsets*, the size of which could be varied independently of changes in the size of the total positive set.

Some evidence from STM recall studies suggests that retrieval from STM is facilitated by the organization present in the to-be-remembered set. Yntema and Trask (1963), in a dichotic listening experiment, demonstrated that *S*s were better in recalling information from two conceptual categories when recall was required by category than by any other recall classification (ears or pairs). Calfee and Peterson (1968) found recall performance was enhanced when lists of eight items were categorized into two sublists. These investigators observed that when two sets, each containing four items, were presented in a grouped or blocked arrangement, two four-item serial position curves were observed, each demonstrating a recency and primacy effect. In the present study, a recognition procedure will be used and *S*s will be required to retrieve items drawn from two subsets (letters and digits). On every trial, however, the total set size will be constant at six items.

---

[1] This study was supported by National Science Foundation Grant GB-16729 awarded to Bowling Green State University and the second author. A portion of the data was reported at the meeting of the Psychonomic Society, San Antonio, November, 1970.

[2] Requests for reprints should be sent to Donald V. DeRosa, Department of Psychology, Bowling Green State University, Bowling Green, Ohio 43403.

In such a situation, it is possible to distinguish at least two strategies which *S*s might employ. A first strategy would involve examining all of the items in the memorized set. Six items are certainly within a normal *S*'s memory span, and it is possible that *S*s will not use the organization present in the input series to determine the presence or absence of the test item in memory. This strategy, termed the *nondirected search*, would involve a search of all memorized items and would result in the absence of any relationship between RT and size of the probed subset. When RT is plotted as a function of size of the probed subset, there should be no relationship between subset size and RT since *S*s would always be searching a set of six items. On the other hand, if *S*s utilized the organization present in the positive set, they might be able to direct and confine their search to only the probed subset. This strategy, called the *directed search*, would be sensitive to variations in subset size, and RT should increase as a function of subset size in the same manner that is typically observed in a single-category condition when set size is varied. Therefore, the directed search predicts increases in RT with subset size comparable to those observed in the single-category case, while the nondirected search would predict the absence of a relationship between RT and subset size.

The ability to use a directed search in a recognition task may be dependent upon at least two factors. First, features of the positive set present at input might influence the use of a particular search strategy. Calfee and Peterson (1968), for example, found recall performance was enhanced when the two categories were presented in a grouped or blocked arrangement in the to-be-remembered series. An additional feature of this investigation will be use of retrieval cues following the presentation of the positive set. These cues will, on some trials, provide information concerning the subset from which the probe will be drawn. The presence of informative cues may facilitate a directed search strategy.

METHOD

Subjects.—The *S*s were 16 male undergraduate volunteers, who were paid for their participation in the experiment. Each *S* served in all conditions of the experiment.

Apparatus.—Each *S* was tested in a sound-deadened isolation chamber. Stimuli were presented on the viewing face of two rear-projection display cells. The cells were enclosed in a wooden chamber, one side of which had a viewing window. All stimuli were projected on the same area of the viewing window by the use of mirrors and one-way glass. Thus, the visual display operated on the same principal as a two-field tachistoscope and allowed the stimuli from two sources to be presented at the same location. The *S*s were seated approximately 3 ft. from the viewing window. Three colored lights situated directly above the viewing window served as cues. Two lights served as informative cues and signaled that the probe would be a letter (red light) or a digit (blue light). A yellow light was used as a noninformative cue and served only to warn *S* that a probe was about to appear. A microphone monitored *S*s' vocal responses, and a voice-operated relay terminated the probe stimulus, advanced a paper tape reader, and activated a Hunter Klockounter which displayed RTs to the nearest .001 sec.

Procedure and experimental conditions.—Each trial consisted of a serial presentation of six stimuli defined as the positive set. The positive set contained both letters and digits chosen from the set of letters A through H and the set of single digits 1 through 8. Each stimulus in the positive set was presented for .4 sec., and there was a .4-sec. interval between the successive stimuli in the set. A cue was shown .4 sec. after the final item in the positive set and remained on for .4 sec. Finally, the probe occurred 2.0 sec. after the cue and remained on until *S* responded.

When the probe appeared, *S* was instructed to respond by saying "yes" if the probe was a member of the positive set and "no," otherwise. Over the entire experiment, 50% of the probes were positive and these stimuli were identical to items in each serial position of the positive set with equal frequency.

Three major variables were manipulated: subset size (number of letters and digits in the positive set), informative versus noninformative cue, and grouped versus nongrouped subsets. Subset size was either two, three, or four items. For example, if the positive set contained two digits, then it would contain four letters. This particular type of trial could be used to examine a subset size of two and a subset size of four as well. Counterbalancing was used to insure that a given subset size (two, three, or four items) was represented by groups of letters and groups of numbers equally often in the same serial position of the positive set.

For each subset size, 50% of the trials were grouped (G); the items from the probed subset were presented contiguously in the positive set. The remaining trials for each subset size were nongrouped ($\bar{G}$); the items from the probed subset were not presented contiguously.

For each subset size in both the G and the $\bar{G}$ conditions, half of the trials had an informative cue (I) preceding the probe; the cue indicated to *S* the conceptual set (either letters or digits) from which the probe would be drawn. The remainder of the trials for each subset size in both the G and the $\bar{G}$ conditions had a noninformative cue ($\bar{I}$) preceding the probe. A base-line condition was also included in which each positive set contained from two to six items which were either all digits or all letters. A noninformative cue (warning signal) preceded the probe.

A total of 576 different experimental trials were divided among 14 lists. Two additional lists of 40 trials each were prepared for the base-line condition. The 16 lists were each presented twice across four 1½-hr. experimental sessions run on 4 consecutive days. A total of 1,312 trials was obtained for each *S*. One day prior to the first experimental session, *Ss* were pretrained and pretested in a 45-min. session. Trials in this period were similar to the experimental trials. The *Ss* whose error rates exceeded 20% in this session were not included in the experimental sessions.

RESULTS AND DISCUSSION

The primary analysis of the data was based on mean RTs for correct responses computed separately for each *S*. The overall error rate was 7.9%, and none of the *Ss* had error rates higher than 13%. The results are presented below in a manner in which first the results for the two conditions of information are shown, and then these data are partitioned, with grouping as a variable.

Directed versus nondirected search.—A major question this investigation focused on was whether *Ss* would utilize the organization present in heterogeneous positive sets to facilitate recognition memory. Organization present in the positive set might allow *S* to form two subsets in memory, and this partitioning could provide the basis for a directed search strategy. The *Ss* might be able to call into active memory only the to-be-probed subset and confine their search to only that subset. On the other hand, only six items were to be retained in memory and this positive set size is well within the span of immediate memory. Correct recognition could still be achieved without resorting to a directed search strategy. A directed search might, however, speed reactions and would result in a linear relationship between subset size and RT. A nondirected search should result in an absence of such a relationship since all six items are examined.

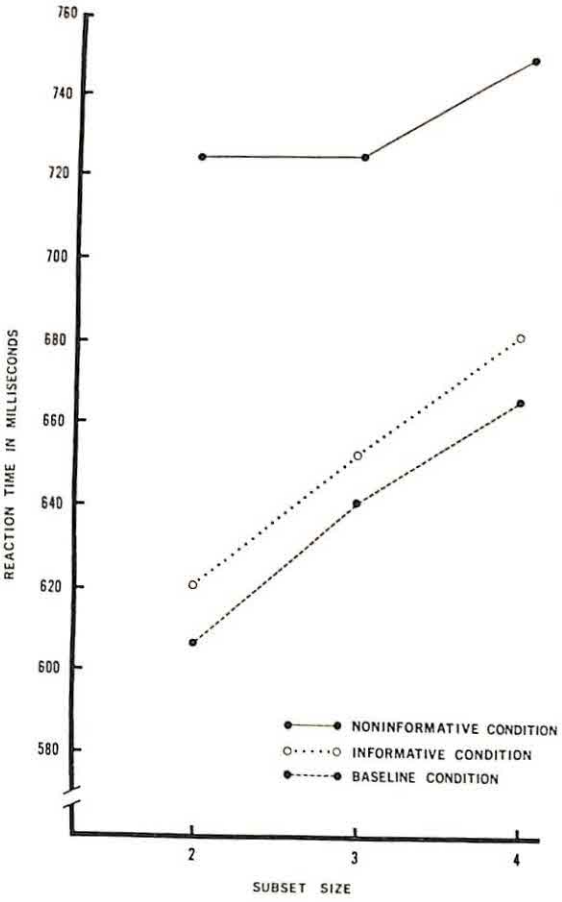


FIG. 1. Mean RT as a function of subset size in the experimental conditions and the base-line condition.

In Fig. 1, RT is plotted as a function of subset size. Three functions are shown in this figure; one represents the base-line condition in which set size was either two, three, or four homogeneous items. The two other functions represent data for the two conditions of information (I vs. $\bar{I}$). For these two functions, RTs are based on the number of items present in the probed subset when sets of six items were shown and probed subset size was either two, three or four items.

Several features of the data in Fig. 1 are noteworthy. First, it is clear that the

TABLE 1

INDIVIDUAL SUBJECT DATA FOR THE INFORMATIVE (I) AND NONINFORMATIVE ($\bar{I}$) CONDITIONS

S	Slope for I (in msec.)	Slope for $\bar{I}$ (in msec.)	Percent slope reduction
1[a]	18.5	1.5	92
2[a]	32.8	1.9	94
3[a]	19.5	3.5	82
4	3.8	−3.1	100
5	24.2	9.2	62
6[a]	9.5	−1.9	100
7[a]	17.0	.8	95
8[a]	22.0	3.2	85
9[a]	38.5	−14.1	100
10	31.9	9.1	72
11[a]	19.0	4.5	76
12	31.1	21.9	30
13	112.2	97.9	13
14	31.6	30.9	2
15	34.2	18.1	47
16	39.0	30.5	22
$\bar{X}$	30.3	13.4	67

[a] These *S*s evidenced a major strategy shift.

base-line or control condition exhibits the typical relationship between set size and RT. RT increases at a rate of approximately 30 msec. per unit increase in set size. For this condition, subset (set) size was highly significant, $F\ (2,\ 30) = 32.29$, $p < .001$. Also, only the linear component of the variance attributable to subset size was significant, $F\ (1,\ 15) = 87.90$, $p < .001$. The linear regression line of best fit for the base-line condition is $RT = 550 + 29.5\ (X)$. A major aspect of Fig. 1 is the fact that the I cue condition produces essentially the identical function as the base-line condition. As was true for the base-line condition, the overall effect of subset size was highly significant, $F\ (2,\ 30) = 24.51$, $p < .001$. Again, only the linear component of the subset size variable was significant, $F\ (1,\ 15) = 26.61$, $p < .001$. The linear regression line for the I function is $RT = 560 + 30.3\ (X)$. The I and base-line conditions do not differ significantly, $F\ (1,\ 15) = 2.55$, $p > .05$; nor is the interaction between these two conditions and subset size significant, $F\ (2,\ 30) = .125$, $p > .05$. When I cues are presented prior to a probe, *S*s appear only to search the probed subset since both the slope and intercept of the I and base-line condition are essentially identical. There is little doubt that *S*s are restricting their search to items in the cued subset and are using a directed search strategy.

The data for the $\bar{I}$ condition present a very different picture. This condition is markedly slower than the I and base-line conditions, and it also produces a marked reduction in slope. The equation which describes this function is $RT = 693 + 13.4\ (X)$. This condition is significantly slower than both the base-line and I conditions, and also a significant interaction is observed between types of information: (I vs. $\bar{I}$) $\times$ Subset Size, $F\ (2,\ 30) = 12.518$, $p < .001$. Although the 13.4-msec. slope is significant, both the linear and quadratic components of the variance attributable to subset size were significant (both ps $< .01$). A further indication of the reliability of the marked difference in slope between the $\bar{I}$ and I conditions is that all 16 *S*s evidenced a slope reduction when $\bar{I}$ cues were present as well as an increase in intercept.

What change in strategy could account for these highly reliable differences between the I and $\bar{I}$ conditions? Inspection of the individual *S* data sheds some light on the reason for these pronounced differences. Shown in Table 1 are individual *S* data comparing performance on the I and $\bar{I}$ conditions. It is rather clear that approximately half of the *S*s (those indicated by a superscript "a") evidenced a major strategy shift. In eight cases, individual slopes in the $\bar{I}$ condition are never greater than 5 msec. per unit increase in set size, and all have a slope reduction of 75% or more. Subject S_4 was not included in this group since his slope was near zero in both the I and $\bar{I}$ conditions. Although the remaining seven *S*s all show a slope reduction, they do so much less than the eight *S*s identified above. It appears that approximately one-half of the *S*s shifted from a directed search to a nondirected search strategy in which they examine all six items of the positive set, while the remaining *S*s (excluding S_4) used some combination of a directed and nondirected strategy, with the majority of these *S*s tending toward a directed strategy.

Grouping.—It was initially proposed that a directed search process might be more probable when subsets were presented in a grouped arrangement in the input series. Grouped sets might result in faster RTs and steeper slopes for functions relating RT to subset size. Figure 2 presents data for the grouped and nongrouped sets. In addition, positive and negative responses are also examined. The left panel of Fig. 2 represents data for trials when I cues preceded probes, while the right panel represents $\bar{I}$ cues trials. The solid data points are means for positive trials. Finally, data points for grouped conditions are connected by solid lines and $\overline{G}$ conditions are represented by broken lines.

For the moment, attention will focus on the data in the left panel of Fig. 2 (informative cue conditions). Grouped sets were reacted to approximately 11 msec. faster than $\overline{G}$ sets. Grouping was a significant variable, F (1, 15) = 11.12, $p < .01$. In addition, the Grouping × Subset Size interaction was also significant, F (2, 30) = 8.30, $p < .01$. Grouped sets resulted in functions having an average slope of 38.0 msec. per unit increase in subset size, while the comparable figure for the $\overline{G}$ functions is 22.5 msec. per unit increase in the subset size. This difference in slope is due entirely to the enhanced reactions to a probed subset of two elements in a grouped positive set. A post hoc comparison of the data for subsets of Size 2 indicated the difference between reactions to G and $\overline{G}$ positive sets was significant, F (1, 15) = 27.54, $p < .01$. Similar comparisons of the data for subsets of Sizes 3 and 4 were not significant. When I cues are present, grouping results in only a slight (11 msec.) decrease in RT, and this decrement is primarily due to the special advantage enjoyed by a probed subset of two elements.

When the results for the $\bar{I}$ condition are examined (right panel of Fig. 2), grouping is shown to have an inhibitory influence on RT. Sets composed of G subsets result in slower RTs (20 msec.) than sets containing $\overline{G}$ subsets. This unexpected result is highly significant, F (1, 15) = 47.31, $p < .001$. An additional indication of the reliability of this finding is the fact that all 16 *S*s were faster when $\overline{G}$ positive sets preceded the probe. Also, the overall reduction in RTs for the $\overline{G}$ sets is essentially the same for all subset sizes. The Grouping × Subset Size interaction was not significant, F (2, 30) = .712, $p > .05$.

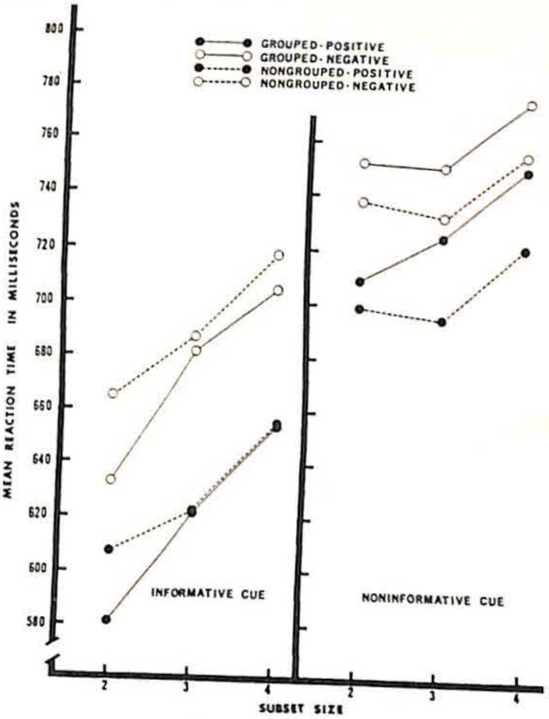


FIG. 2. Mean RT to positive and negative probes for grouped (G) and nongrouped ($\overline{G}$) conditions as a function of size of the positive subset. (The left panel summarizes data for the informative [I] cue condition, and the right panel for the noninformative [$\bar{I}$] cue condition.)

A previous study (Calfee & Peterson, 1968), which utilized grouping in an STM recall paradigm, found enhanced performance when items from the same subsets were presented contiguously in the to-be-recalled series. In a recognition memory RT task, we find that grouping has only a slight facilitative influence when accompanied by I cues and an inhibitory effect when $\bar{I}$ cues are presented.

The results for all variables were the same for positive and negative responses. Positive responses were significantly faster than negative reactions, as is typically the case; however, the type of response (positive or negative) did not interact with any other variable.

Summary and Conclusions

The results of this investigation clearly demonstrate that the organization of a subspan memory set influences short-term recognition memory. When a positive set of six items consisted of two subsets (letters and digits), *S*s utilized this organization in retrieval. If a heterogeneous positive set was followed by an informative retrieval cue, *S*s limited their search to *only* the cued subset. The search of this subset was consistent with the serial exhaustive search model proposed by Sternberg (1966) for homogeneous subspan strings. In addition, the rate of search for I cue trials was identical to the rate observed in a base-line condition (approximately 30 msec. per unit increase in set size). Furthermore, in addition to comparable slopes, the two functions did not differ significantly in average RT. When retrieval cues were $\bar{I}$, a markedly different pattern of behavior was observed. It appeared that approximately one-half of the *S*s shifted strategies to a nondirected search, while the remaining *S*s used a combined directed and nondirected strategy. This may well have varied from trial to trial for these *S*s in the $\bar{I}$ condition.

Presenting the items in a grouped arrangement had only a slight beneficial effect on retrieval. This advantage was limited to I cue conditions. With $\bar{I}$ cues, grouped arrangements of subsets actually significantly retarded RTs and did not change the slope of the function relating RT to subset size. The lack of a facilitative effect is puzzling. It may be associated with a change in motivational level on the part of *S*s experiencing a $\overline{G}$ arrangement of elements in the $\bar{I}$ condition. This condition may have been perceived as the most difficult situation, and *S*s may have expended extra effort in maintaining these items in memory until a probe was shown. This speculation gains some support from evidence obtained by Doll (1971). He demonstrated that RTs were enhanced when monetary rewards were provided for rapid retrieval of information from recent memory. However, within the context of the present study, it is difficult to support the conjecture that motivational level was increased in the $\bar{I}\overline{G}$ conditions without other independent data.

REFERENCES

Calfee, R. C., & Peterson, R. E. Effect of list organization on short-term probe recall. *Journal of Experimental Psychology*, 1968, **78**, 468–474.

DeRosa, D. V. Transformations on sets in short-term memory: Set size reduction by deletion. *Journal of Experimental Psychology*, 1969, **82**, 415–426.

Doll, T. J. Motivation, reaction time, and the contents of active verbal memory. *Journal of Experimental Psychology*, 1971, **87**, 29–36.

Sternberg, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

Yntema, D. B., & Trask, F. P. Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 65–74.

(Received June 29, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 455–457

SHORT REPORTS

SEMANTIC PROCESSING WITH BISENSORY STIMULATION[1]

JOE L. LEWIS[2]

University of Oregon

Dichotic listening experiments have shown that reaction time (RT) to attended words is a function of the semantic similarity of simultaneously presented unattended words. However, D. E. Broadbent and M. Gregory have argued that the ears should be viewed as a simple input channel rather than separate channels. Unattended audio information may be admitted to the processor along with the attended information. An audiovisual presentation was used to examine the level of processing when clearly separate channels were used. Verbal RTs to attended words again indicated that the unattended words were interfering selectively with the attended words. When audio was attended and visual ignored, the ordering of RTs was exactly as found in the dichotic experiments. A significant interaction and different pattern of results when visual was attended and audio ignored indicated that the pattern of interference is not symmetrical for the two modalities.

Proponents of a "filter" theory of selective attention have suggested that unattended stimuli are "blocked" (Broadbent, 1958) or "attenuated" (Treisman, 1969) at a relatively peripheral level. Unfortunately, most of the evidence supporting this hypothesis has been based on the rather insensitive method of recall or on a method requiring *S* to make an overt response to unattended stimuli while simultaneously responding to an attended task. On the other hand, when semantic processing of unattended information has been found, the control over *S*'s attention typically has been inadequate.

A series of dichotic listening experiments (Lewis, 1970) was run to try to provide satisfactory methodological controls. Attention was controlled by (*a*) using a sequence of unrelated words rather than prose, (*b*) using a fast enough presentation rate to prevent *S* from switching attention without losing part of the unattended message, and (*c*) requiring errorless shadowing (repetition while listening) of the attended message. Forgetting was controlled by using interference instead of memory as a measure of processing. The reaction time (RT) to shadow the attended words was measured as a function of their relationship to the simultaneously presented unattended words. It was found that semantic similarity (i.e., synonym pairs) between the attended and unattended words produced interference in repeating the attended words. The unattended words apparently were processed at the semantic level in order to have interfered with the response to the attended words.

However, Broadbent and Gregory (1961) have suggested that the ears may be rather unique as channels. They found that digits presented alternately to the two ears could be recalled as well in the actual order of arrival (alternating ears) as ear by ear. But when stimuli were presented alternatively to the ear and eye, recall was significantly better for the ear then eye than in the actual order of arrival (ear alternating with eye). They suggest that the two ears may be viewed as a single input channel but the eye and ear are separate channels. Consequently, the semantic processing found with dichotic listening may be because all the information from the ears is admitted to the same channel. However, since stimuli may be processed from only one channel at a time, simultaneously presented audiovisual words would not be expected to produce interference from semantic properties. In addition, if the eye and ear are separate channels, the ease of filtering a channel should depend on the spatial contiguity of the channels. Spatial separation of the visual and sound sources should enhance the ability to filter the unwanted information.

Method.—The *S*s were 16 students (6 males and 10 females) from an introductory psychology lecture course at the University of Oregon. All *S*s received credit toward their course grade for participating. The only restriction in soliciting *S*s was that they have 20/20 (corrected) vision and "relatively normal hearing" in both ears.

The same stimulus materials were used as in a previous dichotic listening experiment (Lewis, 1970, Exp. II) to permit a rather direct comparison of the results. Stimuli for the attended messages consisted of 28 lists of five unrelated, single-syllable words approximately matched for frequency of occurrence in English. The unattended messages were constructed using words related to the third word in each attended message. The four kinds of relationship were (*a*) low associative strength nonsynonym associates (low nonsynonyms), (*b*) high associative strength nonsynonym associates (high nonsynonyms), (*c*) low associative strength synonyms (low synonyms), and (*d*) unrelated (controls).

---

[1] This research was supported by the Advanced Research Projects Agency of the United States Department of Defense and was monitored by the United States Air Force Office of Scientific Research under contract F44620-67-C-0099 while the experimenter was at the University of Oregon.

[2] Requests for reprints should be sent to Joe L. Lewis, who is now at the Department of Psychology, University of Washington, Seattle, Washington 98105.

TABLE 1

MEDIAN VERBAL REACTION TIMES

Order and method of presentation	Control		High nonsynonyms		Low nonsynonyms		Low synonyms	
	Audio attended	Visual attended	Audio attended	Visual attended	Audio attended	Visual attended	Audio attended	Visual attended
Audio-visual								
TV	750	715	755	733	760	645	770	690
Headset	610	573	620	575	655	545	685	558
$\bar{X}$	680	644	688	654	708	595	728	624
Visual-audio								
TV	770	618	803	610	770	575	775	600
Headset	525	573	523	593	530	558	530	580
$\bar{X}$	648	596	663	602	650	567	653	590

For the audio-attended condition the 28 attended lists were recorded on magnetic tape at a rate of one word every $\frac{2}{3}$ sec. with 5 sec. between lists. Each of the 28 lists were recorded four times using four random orders. The visual words were constructed by putting each word on a 5 × 7 white card, using Paratipe (Style 11101) pressure lettering. The stimuli were then recorded on videotape using the following procedure. A Uher Model 8000 tape recorder was connected from the audio output via a Grason-Stadler voice-operated relay to the control panel of a Scientific Prototype three-channel tachistoscope. The control was wired so that the display field containing the unattended word was illuminated for 300 msec. when the third audio word began. Background illumination (against a white field) was maintained throughout. Recording on videotape was accomplished by pointing a General Electric closed-circuit television camera into the viewing hood of the tachistoscope and connecting the audio output of the tape recorder to the audio input of a General Electric videotape recorder.

Four blocks of stimuli were constructed using the above procedure. Each block consisted of one set of the 28 audio lists of 1/4 of each of the four types (low nonsynonyms, high nonsynonyms, low synonyms, and controls) of visual (unattended) words. Thus the four treatment conditions were contained in each block yet the attended messages were the same in all four blocks. For example, the critical test word from a given attended list (third word) might be paired with a low nonsynonym in Block 1, a high nonsynonym in Block 2, a low synonym in Block 3, and a control in Block 4.

For the visual-attended condition, the stimuli were recorded on videotape using an analogous procedure such that the five attended words were visual and the unattended word paired with the third visual word was audio.

Twenty lists of training stimuli were also constructed for each viewing condition.

No equipment was available to measure the accuracy of synchronization between the audiovisual pairs. Two judges monitored all the stimuli and were unable to detect any asynchrony.

The Ss were assigned randomly to one of four groups to satisfy the constraints of a 2 × 2 matrix. Half of the Ss attended to visual words during the first half of the session and were switched to audio words during the last half of the session. The other half of the Ss received the reversed order of presentation. Half of each of the above two groups were assigned to hear all audio words over a low-impedance headset (separate source), and the other half were assigned to receive the audio words from the television speaker (same source). All Ss received messages from all four treatment conditions (e.g., low synonyms, etc.). Thus a 4 (Treatment) × 2 (Modality Attended) × 2 (Order of Attending) × 2 (Separation) design was established. The S was given a brief description of the apparatus and told that his task was to listen to (or view) the audio (visual) words and to repeat (shadow) each word as soon as he heard it. For the audio-attended presentation, S was told to watch the TV but to ignore any visual words. For the visual-attended presentation, S was told to ignore any audio stimuli. The S was trained to shadow to the criterion of 5 errorless lists. Training lists were repeated if more than 20 were required. The four blocks of treatment lists were then presented. A 2-min. rest was given following training and between each block of treatment lists. A 5-sec. intertrial interval otherwise separated each list. A 5-min. rest followed the fourth block and the entire procedure was repeated except that S attended to the modality which was unattended during the first half of the session. The S spoke into a microphone and all responses were recorded on Channel 1 of a magnetic tape. The audio stimuli were recorded simultaneously on Channel 2 of the same tape.

Verbal RTs were obtained by later playback of the recording of the audio stimuli with S's respons The output of the recorder was connected so that each channel fed into a different channel of a Gils analog pen recorder. A divide circuit was used t invert the negative part of the sine wave so that positive tracing was produced. The paper speed of the pen recorder was set to record 10 msec/m The RTs were obtained by measuring the distan between the rise produced by the audio stimuli an the rise produced by S's response.

Results and discussion.—Median RTs to the thir position attended words are shown in Table 1 Analysis of variance revealed significant effects f treatments, F (3, 36) = 3.52; separation, F (1, 1 = 7.80; Treatment × Modality Attended inter tion, $F(3, 36) = 2.89$, and Order × Modality Attended × Separation interaction, $F(1, 12) = 5.56$; in all cases, $p < .05$.

The separation main effect was produced by *Ss* responding faster when presentation was through the headset than when presentation was through the TV speaker. This was likely due to the masking of the TV speaker sound by *S*'s own responding. The critical Treatments × Separation interaction was not significant, $F(3, 36) = .78$. The hypothesis that the spatial separation of the visual and sound sources will enhance filtering was not supported. However, the significant Order × Modality Attended × Separation interaction does raise some question about the additivity of the effects (see Sternberg, 1969). Apparently, transfer of attention from one modality to another during the session may have adverse effects on attention to the second modality.

The main effect for treatments was significant, indicating that even when the channels are separated by using the ear and the eye, interference as a function of semantic properties of the messages occurs. However, the Treatments × Modality Attended interaction indicates that the effects for the two modalities are not symmetrical for the senses. The RTs from the first half of the session along with the comparable RTs obtained with dichotic listening (Lewis, 1970) are shown in Table 2. Only the data from the first half of the session are shown because the first half of the session in the present experiment corresponds to the type and number of trials from the dichotic experiment.

TABLE 2

MEDIAN VERBAL REACTION TIMES FROM FIRST HALF OF SESSION AND FROM AUDIO-AUDIO EXPERIMENT

	Control	High non-synonyms	Low non-synonyms	Low synonyms
Visual attended	596	602	567	590
Audio attended	680	688	708	728
Dichotic (Lewis, 1970)	687	690	718	726

It is clear from Table 2 that when the attended message is audio the interfering effects are very similar regardless of the modality of the unattended message. The ordering of the RTs is the same in both cases. (It should be noted that the absolute values given in Table 2 have no special significance since the headset and TV medians are pooled.) The similarity between the dichotic experiment and the audio-attended condition is inconsistent with Broadbent and Gregory's (1961) suggestion that the ears are unique as channels. There is no indication that unattended visual information must wait until the channel is free while unattended audio information is admitted to the processor along with the attended audio information. A somewhat different picture emerges from the visual-attended, audio-unattended condition. The significant Treatment × Modality Attended interaction does indicate that the processing is different at some stage, depending on the modality of the input. The only effect significantly different from the controls is the low nonsynonyms. Analysis of this group indicates that it contains a large number of pairs which are frequently sequential in English (e.g., STREETCAR; see Lewis, 1970). When visual is attended, these appear to have a facilitation effect.

Another important result emerged between the recall results in the dichotic experiment and the present experiment. With dichotic listening, recall of unattended information is very low or zero. In the present experiment, recall was very high with both modalities of attention. Several *Ss* volunteered information about the content of the unattended message. Interrogation of *Ss* revealed good recall of the unattended content. Although Broadbent and Gregory (1961) do not discuss the results of recall between their two experiments, a similar effect seems to be apparent in their data. The proportion of lists recalled ear by ear in the dichotic experiment was .28, while the corresponding result from the ear-eye experiment was .59.

REFERENCES

BROADBENT, D. E. *Perception and communication.* New York: Pergamon, 1958

BROADBENT, D. E., & GREGORY, M. On the recall of stimuli presented alternately to two sense organs. *Quarterly Journal of Experimental Psychology*, 1961, **13**, 103–110.

LEWIS, J. L. Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology*, 1970, **85**, 225–228.

STERNBERG, S. The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, 1969, **30**, 276–315.

TREISMAN, A. M. Strategies and models of selective attention. *Psychological Review*, 1969, **76**, 282–299.

(Received November 27, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 458–460

PARTIAL DELAY OF REWARD IN THE DOUBLE ALLEYWAY

JOSEPH A. SGRO,[1] NEIL H. COHN, AND STEPHEN D. DUDLEY

Virginia Polytechnic Institute and State University

Two groups of rats ($N = 40$) were given 54 trials with either continuous reward or partial delay of reward in Goal-box 1 of a double alleyway. All *Ss* were than shifted to partial delay for 36 trials. During preshift, the partial group showed slower Alley 2 speeds following delay than following no delay. During postshift, both groups showed depressed Alley 2 speeds following delay trials in Goal-box 1, the effect being greater for the group which received prior exposure to continuous training. The results were discussed in relation to theoretical and empirical considerations of nonreward and delay in the double alleyway.

In a typical experiment demonstrating the motivational effects of nonreward, rats are trained with continuous reinforcement in the first goal box of a double alleyway and then shifted to partial reinforcement. The result is faster second alleyway speeds following nonreward than following reward, a finding commonly termed the "frustration effect" (FE). A recent study by Sgro, Glotfelty, and Moore (1970) attempted to demonstrate that shift from immediate reward to partial delay of Goal-box 1 reward operated in a similar fashion. Instead of delay-facilitated Alley 2 speeds, it was found that speeds following delay of Goal-box 1 reward were *slower* than speeds following no delay.

Both of the above phenomena have been interpreted within the framework of frustration theory. While the FE has been attributed to the motivational properties of primary frustration (R_F) generated in Goal-box 1 (Amsel, 1958), slower Alley 2 speeds following delay of reward in Goal-box 1 have been explained on the basis of competing responses which arise from delay-engendered R_F. The responses become conditioned to stimuli in Goal-box 1, thereby interfering with *Ss* leaving Goal-box 1 when provided access to Alley 2 (Sgro, Showalter, & Cohn, 1971).

While the FE has been shown to develop after a series of continuously reinforced trials (Amsel, 1958), it also has been reported under conditions in which rats experienced partial reinforcement in Goal-box 1 throughout the experiment (Wagner, 1959). With particular attention to the studies by Sgro et al. (1970), the question remains whether depressed Alley 2 speeds following delay of Goal-box 1 reward are contingent upon prior exposure to continuous immediate reward or whether slower Alley 2 speeds might be exhibited under extended partial delay of reward conditions. To answer this question, one group was given 90 trials with partial delay of Goal-box 1 reward. A second group received 54 trials of immediate reward and then was shifted to partial delay of reward for 36 trials.

Method.—The *Ss* were 40 female hooded rats obtained from the laboratory colony of the Department of Psychology at Virginia Polytechnic Institute. The *Ss* were 90–100 days old at the beginning of the experiment.

The apparatus was an L-shaped double alleyway identical to that employed by Sgro, Glotfelty, and Podlesni (1969). The apparatus consisted of a first start box, a first alleyway, a first goal box, a second alleyway, and a second goal box. The first goal box also served as the start box for the second alleyway. A horizontal sliding door was located 7.62 cm. from the guillotine door which separated the first goal box from the second alleyway.

Running times for Alley 2 were measured from the opening of the Goal-box 1 door leading into Alley 2 until the interruption of a photobeam located 190.50 cm. distant.

The *Ss* were randomly assigned to two equal groups and received either continuous reward (Group C) or partial delay of reward (Group P) in Goal-box 1. After 54 trials, Group C was shifted to the same partial delay of reward schedule in Goal-box 1 as Group P for an additional 36 trials. The pattern of immediately rewarded (R) and delayed (D) trials in Goal-box 1 for Group P during preshift was: RRD, DRD, RDR, DDR, DRR and RDD. This pattern was repeated three times over the preshift and twice over the postshift phase.

Twelve days prior to the first preshift trial, all *Ss* were placed on a 22½-hr. food deprivation schedule and handled approximately 2 min. daily. Upon completion of each daily session, all *Ss* were allowed 1½ hr. access to food in the home cage. On Days 11 and 12, each *S* was allowed to explore the entire apparatus for 2 min. in addition to being given scattered 45-mg. Noyes food pellets in a neutral holding cage. On Preshift Day 1, *S* was given one trial; on Day 2, two trials; and three trials daily thereafter.

The intertrial interval was approximately 20–30 sec. During the intertrial interval, *S* was placed in a neutral holding cage. The typical procedure on a trial consisted of removing *S* from its home cage and placing it into the start box. Five seconds later, *E* manually opened the guillotine door, allowing *S* to traverse Alley 1. As soon as *S* entered Goal-box 1, the guillotine door was closed to prevent retracing and *S* was permitted to eat four 45-mg. Noyes food pellets. On delay trials, the same procedure was used except that the horizontal sliding delay door remained closed for 15 sec. after *S* entered Goal-box 1.

---

[1] Requests for reprints should be sent to Joseph A. Sgro, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

Upon completion of the eating and as soon as *S* was oriented to the guillotine door leading to Alley 2, the door was manually raised and *S* was allowed to traverse Alley 2. Four 45-mg. Noyes food pellets were available in Goal-box 2 throughout the experiment. After completion of the daily trials, *S* was not given the daily ration in the home cages until 15 min. had elapsed.

Results and discussion.—The mean Alley 2 running speeds (m/sec) as a function of trial blocks are presented in Fig. 1. Observation of Group P reveals that the separation between speeds following R and D trials in Goal-box 1 occurred on Trial Block 5 and continued throughout the remainder of the study. Analysis of variance performed over the nine preshift trial blocks for Group P yielded a significant blocks effect as well as a significant Blocks × Goal-box 1 Delay Condition interaction, F (8, 152) = 46.20, $p < .01$, and F (3, 152) = 2.32, $p < .05$, respectively. Analysis of variance performed over Trial Blocks 5–9 revealed that Alley 2 speeds following delay were significantly slower than speeds following no delay F (1, 19) = 12.83, $p < .01$.

Inspection of the postshift phase within Fig. 1 shows that upon the introduction of partial delay of Goal-box 1 reward, Alley 2 speeds following D trials for Group C revealed an immediate decline to a level below Group P. Observation of the mean Alley 2 speed for Group P on each of the three D trials within Trial Block 10 were .69, .77 and .72, respectively. Correspondingly, the mean Alley 2 speeds for Group C on the same trials were .54, .69, and .64. It is interesting to note, however, that Alley 2 speeds following R trials were not different for the two groups. Analysis of variance performed over Trial Blocks 10–15 indicated that the Goal-box 1 delay-condition variable was significant, F (1, 38) = 104.83, $p < .01$, as was the Goal-box 1 Delay Condition × Preshift Group interaction, F (1, 38) = 5.39, $p < .05$.

During preshift, the presence of slower Alley 2 speeds following delay than following no delay for Group P indicate that a shift from immediate to partial delay of Goal-box 1 reward is not a necessary condition for the effect to occur. It is interesting to note that in the Sgro et al. (1971) study, a group was run with partial delay training from the outset. Since the immediate versus delay trials were not reported separately, the data was reanalyzed for purposes of comparison. It was found that while Alley 2 speeds were observed to be slower following delay than following no delay, the difference was not statistically significant. The failure to achieve statistical significance within Sgro et al.'s (1971) group could have been attributed to differences in situational and procedural variables and less sensitivity than in the present study. The postshift data is in agreement with the findings of Sgro et al. (1970) and indicate that prior exposure to a series of immediately rewarded trials in Goal-box 1 serves to accentuate the degree of depression in Alley 2 speeds following D trials in Goal-box 1.

Considering the comparison of partial reward and partial delay of reward in the double alleyway, it has been assumed (Amsel, 1958) that with FE the size of R_F is dependent upon the extent of reward expectancy. With regard to the relationship between number of training trials and the FE, Yelen (1969) has shown that the FE is greater following 60 continuously reinforced trials than following either 12 or 36 trials. To the extent that delay of reward constitutes a frustrative event, it should be expected that since prior exposure to continuous immediate reward would serve to increase reward expectancy, delay-engendered R_F should be greater for Group C than Group P. It has been suggested (Sgro et al., 1971) that since reward always follows the delay interval, it would be expected that competing responses generated by R_F would become conditioned to stimuli in Goal-box 1. The suggestion (Sgro et al., 1971) that competing responses generated by R_F became conditioned to stimuli in Goal-box 1 as a consequence of delay always being followed by reward is questionable in light of the present data. Since the first postshift D trial by Group C resulted in slower Alley 2 speeds than Group P, it would appear that a more reasonable interpretation is that the initial effect of shift to delay was to produce competing emotional responses that interfered with leaving Goal-box 1. Whether conditioning did occur during postshift is problematic, however, since conditioning could have occurred only to be masked by the continual generation of competing responses on D trials. Within the above reasoning, it would be expected that upon shift to partial delay, Group C should show slower Alley 2 speeds on D trials than Group P.

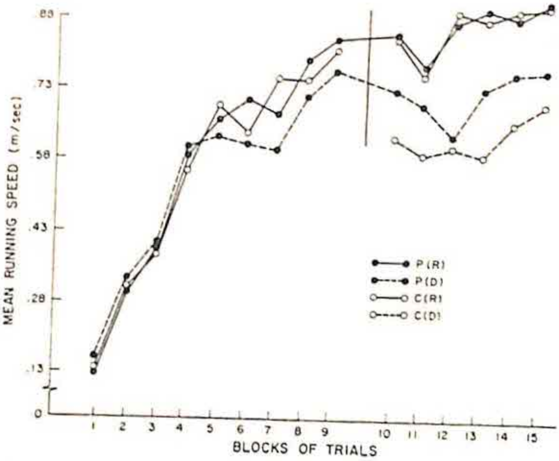


FIG. 1. Mean Alley 2 running speeds as a function of trial blocks. (Trial Blocks 1–9 for Group C are plotted in blocks of six trials. All other trial blocks for both groups are plotted in blocks of three trials.)

A final point deserves mentioning. Wagner (1959) has shown that no difference exists in the magnitude of FE between a group trained with partial reinforcement in Goal-box 1 and one trained with continuous and shifted to partial reinforcement. Observation of Wagner's data, however, reveals that lack of a significant Group × FE interaction might be attributed to a "ceiling effect" on Alley 2 speeds following nonreward in Goal-box 1. According to frustration theory, it would be predicted that the greater the number of continuously reinforced trials, the greater the FE.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

SGRO, J. A., GLOTFELTY, R. A., & MOORE, D. B. Delay of reward in the double alleyway: A within-subjects versus between-groups comparison. *Journal of Experimental Psychology*, 1970, **84**, 82–87.

SGRO, J. A., GLOTFELTY, R. A., & PODLESNI, J. A. Contrast effects and delay of reward in the double alleyway. *Psychonomic Science*, 1969, **16**, 29–31.

SGRO, J. A., SHOWALTER, J. R., & COHN, N. H. Frustration effect following training with continuous and partial delay of reward. *Journal of Experimental Psychology*, 1971, **87**, 320–325.

WAGNER, A. R. The role of reinforcement and nonreinforcement in an "apparent frustration effect." *Journal of Experimental Psychology* 1959, **57**, 130–136.

YELEN, D. Magnitude of the frustration effect and number of training trials. *Psychonomic Science*, 1969, **15**, 137–138.

(Received December 6, 1971)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 460–462

SERIAL POSITION EFFECTS IN PROBE RECALL: EFFECT OF REHEARSAL ON REACTION TIME[1]

JOHN G. SEAMON[2]

New York University

Serial position effects in latency data were investigated in a recall paradigm. The *S*s learned a serial list comprised of four categories of three words each, with memory subsequently probed. Results showed that (*a*) when rehearsal processes are uncontrolled, significant primacy and recency effects occur within and between categories, (*b*) additional practice on selected items during learning produces significantly faster reaction times during testing, and (*c*) equivalent rehearsal for all items eliminates the significant primacy and recency effect between categories, but not within categories. Serial position effects are discussed in terms of differential rehearsal and memory trace accessibility.

In a free-recall task, *S* is presented with a sequential list of words and is then asked to recall as many words as possible in any desired order. The results typically show that words presented early or late in the list are recalled with higher probability than words from the middle of the list. Rundus and Atkinson (1970) attribute the serial position effect to the rehearsal processes employed by *S* during list presentation. They had *S*s freely rehearse words out loud while a list was presented for study. It was found that items from the beginning of the list were rehearsed more frequently than items from all other positions, and items late in the list had the highest probability of being rehearsed at the time of recall. Rehearsal may thus be seen as a process for maintaining information in a readily available state or as a means of strengthening the memory trace.

Similar serial position effects have been found in short-term memory studies of item recall (Norman & Wickelgren, 1969), position recall (Moss & Sharac, 1970), and item recognition (Corballis, 1967; Morin, DeRosa, & Stultz, 1967; Norman & Wickelgren, 1969), with reaction time (RT) as the dependent variable. In general, RT is fastest for the most recently presented items and longest for items presented in the middle of the sequence. It may be that rehearsal processes are responsible for the serial position effects in latency data as in free recall. To test this hypothesis, *S*s were given lists of words to learn with memory subsequently probed in an item recall paradigm. The first study permitted free rehearsal during learning, the second provided additional rehearsal for selected items, and the third equated rehearsal for all items. Latency differences between serial positions, if due solely to rehearsal processes, should be present in the first study, altered in the second, and not found in the third.

Method.—In the first experiment, 10 introductory psychology students from the University of Massachusetts served individually as *S*s to fulfill a course requirement. Stimuli and test trials were presented on slides with a Kodak Carousel RA 950 projector. The RTs on test trials were obtained in milliseconds from a Hunter timer in concert with a Gerbrands voice key, Lafayette tachistoscope, and the Kodak projector. The closure of a manual switch simultaneously opened the projector shutter, started the timer, and activated the voice key. A verbal re-

---

[1] This research was conducted at the University of Massachusetts and prepared during a postdoctoral fellowship at New York University.

[2] Requests for reprints should be sent to John G. Seamon, who is now at the Department of Psychology, Wesleyan University, Middletown, Connecticut 06457.

TABLE 1

MEAN RESPONSE LATENCY (IN MSEC.) FOR SERIAL POSITION MEMORY PROBES

Experimental conditions	Serial position												Overall errors
	1	2	3	4	5	6	7	8	9	10	11	12	
Exp. I													
Original	1090	1353	1416	1217	1539	1453	1505	1628	1532	1586	1665	1332	2.57%
Replication	1137	1432	1333	1541	1694	1540	1404	1580	1498	1695	1718	1391	2.11%
Exp. II													
Control group	1291	1517	1503	1374	1414	1420	1476	1830	1677	1501	1627	1275	3.64%
Experimental group	1221	1462	1333	1409	1471	1423	1191	1412	1394	1372	1487	1233	2.18%
Exp. III													
First 10 trials	1422	1578	1353	1510	1534	1524	1443	1524	1470	1506	1591	1417	3.05%
All 25 trials	1339	1476	1287	1376	1490	1414	1375	1487	1398	1434	1528	1338	1.10%

sponse from *S* into the voice key terminated the system.

Using the serial recall method, each *S* learned a serial list comprised of four categories of three words each, viz., TREE-BUSH-FERN-PONY-GOAT-LAMB-SHOE-BELT-VEST-BONE-SKIN-HAIR. A categorized list was used to observe if serial position effects would be found between and within the category units. The *S*s were run individually, with each word shown for 5 sec. followed by a recall interval of 30 sec. after the last slide. The *S* was instructed to recall the words during this interval by writing them in their correct order on a sheet of paper which had the appropriate number of blanks provided. Only the final serial arrangement of the words was important. The *S* could fill in the response sheet in any desired temporal order. A new answer sheet was used on each trial with no feedback given. A criterion of two successive, perfect recall trials was employed. Use of well-learned material insured that all memory traces would be available.

Following completion of the serial recall task, *S* was presented with memory probes for each word in the list with RT recorded. A memory probe consisted of the simultaneous presentation of a category name and an item position cue, e.g., $\begin{bmatrix}\text{BODY}\\ \_\ \text{x}\ \_\end{bmatrix}$. The *S* was required to fill in the word that occurred in the position signified by the x, in the category named, as quickly and as correctly as possible by making the response verbally into the voice key. Approximately 3 sec. elapsed between memory probes. One presentation of the individual probes for all list words constituted a trial and was followed by a 30-sec. rest period. A total of 10 trials was used, with the memory probes presented in a random order on each trial.

In the second experiment, 16 students served as *S*s. The *S*s were divided equally into a control and an experimental group. The procedure for the control *S*s was the same as that used in the preceding experiment. Experimental *S*s differed by having two presentations and recalls of the words of the third category in the list prior to learning the whole list. After the experimental *S*s viewed and recalled the three-word category twice, the list-learning and memory-probe procedures were identical for both groups. The same words used previously were employed, with the order of the categories and the items within the categories randomly determined for each *S*.

Six students participated as *S*s in the third experiment. Unlike the previous experiments in which one word at a time was presented for learning, this experiment presented all the words of a given category simultaneously. The words of a category were displayed horizontally on the projection screen for 18 sec. While the words were being shown, *S*s were required to rehearse them by orally reciting the entire group sequentially between auditory clicks that were provided by a solenoid switch and two Hunter timers. The clicks were temporally spaced to provide .4 sec. to rehearse each word displayed. An auditory click every 1.2 sec. for 18 sec. enabled each word to be rehearsed 15 times. The recall and memory-probe procedures were the same as those used previously, with the number of memory-probe trials increased from 10 to 25 to obtain more observations as only six *S*s were used. The same words were employed, with a random order of categories and items within categories for each *S*.

Results and discussion.—The RT data are essentially error free as shown by the mean error rate for each condition in Table 1. In each experiment, a mean RT was determined for each *S* for each serial position, with a mean then obtained for each serial position over all *S*s. The RT data for all experiments are presented in Table 1.

In the first study, RT is clearly a function of the item position within a category and the category position within the list. The individual *S* means were used in separate analyses of category and item position. The analysis of variance for RT as a function of category position was significant, $F\ (3, 27) = 11.77$, as was the analysis of variance for RT over item position, $F\ (2, 18) = 13.57$, both $ps < .001$. In general, there were primacy and recency effects within and between categories.

Serial position and stimulus word were confounded in this experiment as each *S* learned the same list. Although unlikely, it may be that the RT differences reflect recall differences for the individual words. A replication experiment was performed to test this possibility and the reliability of the results. Ten additional *S*s from the same pool were used. The words learned by *S*s were the same as those used previously, but the serial order of the categories relearning the first VD list, beginning with a test trial, or a modified free-recall test in which *S*s were shown one member of a VD pair and asked to produce the item or items that had been paired with it. The final variable was retention interval. The retention test was administered either 10 min. or 25 min. after first-list acquisition. This variable was included in view of Eschenbrenner's (1969) demonstration that RI effects were more dramatic on delayed VD tests. Thus, there were 12 major conditions derived from the factorial combination of the three paradigms, the two types of retention test, and the two retention intervals.

The lists were constructed in a rather complex fashion, but in essence there were four different 16-pair lists that were used equally often as the first list *S*s learned. The lists were constructed from a pool of 80 common English words. From this pool of 80 words, 16 were selected randomly and without replacement to serve as correct alternatives in the first VD list. An additional 16 words were selected to serve as incorrect alternatives and were paired randomly with the members of the first set. A third set of 16 words was selected to serve as new words in List 2, new correct words in the W_1-R_2 paradigm, or new incorrect words in the W_2-R_1 paradigm. Of the remaining 32 words, 16 were selected randomly to serve as correct alternatives and the remaining 16 served as incorrect alternatives in a second version of List 1. Thus, there were two VD lists comprised of different words, with one serving as the first list for half of the *S*s and the other serving as the first list for the remaining *S*s. Finally, for each of these two distinct lists, the sets of correct and incorrect items were reversed for half of the *S*s.

The VD training began with a study trial during which the 16 pairs were presented successively on a Stowe memory drum at a 2-sec. rate. The correct member of each pair was underlined, with the correct alternative appearing on the right for half of the pairs and on the left for the other half. Each study trial was followed by a test trial, during which each pair appeared for 2 sec. with neither member underlined. There were three random orders of pairs for the study trials and three random orders for test trials. First-list acquisition was continued until a criterion of one perfect trial was recorded.

For the experimental groups, second-list acquisition began after a brief instruction. The instruction included appropriate information concerning the specific paradigm; i.e., either all old incorrect or all old correct items would be retained. The *S*s were given six study and test trials on List 2. The remainder of the retention intervals for experimental groups (25-min. interval) and the entire retention intervals for the single-list groups were spent solving arithmetic problems.

The retention test involved either relearning the VD list (List 1) or associative recall. Relearning began with a test trial and continued until a criterion of two successive perfect trials was reached. The associative-recall test consisted of the presentation of either all correct alternatives (for one-half of the *S*s) or all incorrect alternatives (for the remaining one-half of the *S*s), with *S*s asked to supply the missing alternatives. The *S*s in W_1-R_2 groups provided with incorrect alternatives and *S*s in W_2-R_1 groups provided with correct alternatives were asked to recall both List 1 and List 2 associations. The *S*s in W_1-R_2 groups provided with List 1 correct alternatives and *S*s in W_2-R_1 groups provided with List 1 incorrect alternatives had only one association for each cue. The associative recall was unpaced.

The 12 major conditions and the four distinct first lists combined to yield 48 subgroups. In addition, each of the associative-recall groups was further divided depending upon whether associative recall was cued by the correct VD alternatives or the incorrect VD alternatives. A total of 192 introductory psychology students served in the experiment, with 4 *S*s assigned at random (based upon their order of appearance at the laboratory) to each of the 48 subgroups. There were 16 *S*s in each of the 12 major conditions. Sixteen of the original 192 *S*s failed to reach criterion on the first list within 20 trials. These *S*s were replaced by students from the same *S* pool. The discarded *S*s were distributed evenly across conditions with the exception that 8 were in the single-list control condition, 7 were in the W_2-R_1 condition, and only 1 was in the W_1-R_2 condition. Essentially, all *S*s learned the same first list, and the selection factor probably served to reduce the initial random variability. However, the possibility remains that control and W_2-R_1 conditions had faster learning *S*s than W_1-R_2.

Results and discussion.—The initial comparison was based upon a trials-to-criterion measure for first-list acquisition. A 4 × 3 × 2 × 2 analysis of variance was used, although 3 of the 4 variables were manipulated subsequent to List 1 learning. There was no effect for variation in List 1 materials and this variable did not interact with any other variable. The only significant effect in comparisons of first-list acquisition involved an unexplained Paradigm × Type of Retention Test interaction, two variables not manipulated until a later stage in the experiment, $F\ (2, 144) = 3.27$, $p < .05$. The average number of trials to criterion for groups which later had the associative-recall test was 6.7, 7.4, and 9.0 for the single-list control, the W_2-R_1 paradigm, and the W_1-R_2 paradigm, respectively. The VD test groups averaged 8.3, 6.7, and 6.5 trials for control, W_2-R_1, and W_1-R_2, respectively.

A 4 × 2 × 2 × 2 analysis of variance was used to compare the experimental groups on total errors during the six List 2 trials. This analysis revealed only one significant effect. There were more errors made by *S*s who learned a second list conforming to a W_1-R_2 paradigm than to a W_2-R_1 paradigm, $F\ (1, 96) = 70.59$, $p < .01$, a result in general agreement with previous reports (e.g., Underwood, Jesse, & Ekstrand, 1964).

The data of major interest concerned performance on the retention tests. The level of performance on the VD test trial was very high with little variance in contrast to the performance and variance associated with the associative-recall retention test. In view of the extreme heterogeneity of variance which existed across these recall measures, the VD and associative-recall data were analyzed separately. The variance associated with performance on the VD test trial was 1.51, and the variance associated with the associative-recall measure was 11.96, $F\ (95, 95) = 7.92,\ p < .01$.

Relearning of the first VD list began with a test trial. Trials to relearn and errors on the first trial served as dependent variables in separate analyses of variance. In both analyses, the only significant effect involved differences across paradigms. The mean trials to criterion (including the two criterial trials) for the single-list control groups was 3.3; for the W_2-R_1 groups the mean was 2.6; and for W_1-R_2 groups the mean was 4.9, $F\ (2, 72) = 10.87,\ p < .01$. The mean errors on the first test trial for the control groups was .81; for the W_2-R_1 groups the mean was .25; and for the W_1-R_2 groups the mean was 1.75, $F\ (2, 72) = 15.97,\ p < .01$. In view of the large number of perfect scores (e.g., zero errors), both analyses were repeated following a square-root transformation in the scale of measurement. This transformation did not result in any change in the significance levels of the results.

It was noted earlier that frequency theory predicted RI for W_1-R_2 groups and retroactive facilitation for W_2-R_1 groups. The group differences are consistent with these predictions, but to provide a direct test of these predictions multiple t tests were performed. Significant RI was present with trials to relearn, $t\ (62) = 3.25,\ p < .01$, and with errors on the first VD test trial, $t\ (62) = 3.48,\ p < .01$. The evidence for retroactive facilitation was not as clear, as only the comparison involving errors on the first VD test trial was significant, $t\ (62) = 2.07,\ p < .05$. Again, a transformation in the scale of measurement did not result in any changes in the reported significance levels.

Associative recall was scored in two ways: A stringent recall score treated recall as correct only if the recalled item was paired with its appropriate alternative, and a lenient score was based upon total number of items recalled regardless of pairing. A $4 \times 3 \times 2 \times 2$ analysis of variance (List × Paradigm × Retention Interval × Correct vs. Incorrect Item as Cue) revealed only one significant effect for the stringent recall score. There was a significant interaction between the variables of paradigm and retention interval, $F\ (2, 48) = 3.55,\ p < .05$. The mean numbers of words recalled for the 10-min. delay groups were 5.2 for the control, 3.9 for W_2-R_1, and 6.1 for W_1-R_2. The corresponding values for the 25-min. delay groups were 6.4 for the control, 5.8 for W_2-R_1, and 3.8 for W_1-R_2. In view of this interaction, a series of t tests were performed comparing each experimental group against its appropriate control. The only significant difference appeared after a 25-min. interval as the control was superior to W_1-R_2, $t\ (30) = 2.27,\ p < .05$.

A similar analysis of variance involving lenient recall scores revealed that recall was higher when *S*s were given the incorrect alternatives and asked to produce the correct items ($\bar{X} = 6.75$) than when the reverse conditions applied ($\bar{X} = 5.29$), $F\ (1, 48) = 4.15,\ p < .05$. In addition, there was a List × Retention Interval interaction, $F\ (3, 48) = 2.85,\ p < .05$. For reasons which are not understood clearly, there was better recall at the 10-min. interval, compared to the 25-min. interval, for two of the list conditions and better recall at the 25-min. interval for the other two list conditions.

The present study failed to replicate the evidence for associative interference in modified free recall following VD acquisition for a W_2-R_1 paradigm (Eschenbrenner & Kausler, 1968). For the W_1-R_2 paradigm, there was evidence for associative interference only at the 25-min. retention interval. Since there were numerous procedural differences between the present study and the earlier reports, reasons for the failure to replicate cannot be specified. Although it is difficult to present decisive evidence for a lack of "covariation," the important aspect of these data is that variations in RI as a function of paradigm demonstrated in a VD retention test and relearning are not accompanied by similar variations in RI for associative recall. Since such covariation appears to be at the heart of claims for the necessity of involving incidental learning processes in VD, the present results are considered to be discrepant with such claims. In summary, the VD data seem to square reasonably well with frequency theory, and it does not appear to be necessary to implicate "associative interference mechanisms" in the explanation of these retroactive effects in VD.

REFERENCES

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

Eschenbrenner, A. J., Jr. Retroactive and proactive inhibition in verbal discrimination learning. *Journal of Experimental Psychology*, 1969, **81**, 576–583.

Eschenbrenner, A. J., Jr., & Kausler, D. H. Unlearning of List 1 wrong items in verbal discrimination transfer. *Journal of Experimental Psychology*, 1968, **78**, 696–698.

Kausler, D. H., Fulkerson, F. E., & Eschenbrenner, A. J., Jr. Unlearning of List 1 right items in verbal discrimination transfer. *Journal of Experimental Psychology*, 1967, **75**, 379–385.

Underwood, B. J., Jesse, F., & Ekstrand, B. R. Knowledge of rights and wrongs in verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 183–186.

(Received February 11, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 466–467

EFFECTS OF STIMULUS AND RESPONSE PATTERNS ON CHOICE REACTION TIME [1]

CHARLES P. WHITMAN AND E. SCOTT GELLER [2]

Virginia Polytechnic Institute and State University

The effects of stimulus and response patterns on reaction latency to identify the outcome of a stimulus prediction were studied. Prior to each stimulus presentation, *S*s predicted which of two stimuli would occur and then executed one response if their prediction had been correct or an alternative response if their prediction had been incorrect. Thus, four second-order patterns were possible: stimulus repetition–response repetition, stimulus repetition–response alternation, stimulus alternation–response repetition, and stimulus alternation–response alternation. Choice latencies were shorter when both the stimulus and the response alternated or repeated than when either repeated while the other alternated. The results supported the notion that *S*s hesitate to repeat responses to different stimuli or to make different responses to the same stimulus.

Of the several investigators who have examined sequential effects in a choice reaction time (RT) paradigm, some have observed negative recency or faster reactions to alternated stimuli (e.g., Hale, 1967), others have found positive recency or faster reactions to repeated stimuli (e.g., Remington, 1969), and some have observed interactions between probability and recency effects. Specifically, Kornblum (1969) demonstrated positive recency when stimulus repetitions were more probable than alternations and negative recency when alternations were more probable than repetitions. Additionally, recent two-choice RT experiments demonstrated positive recency to the more probable stimulus and negative recency to the less probable stimulus (e.g., Whitman & Geller, 1972). When alternations and repetitions have been equiprobable, positive recency generally has been observed.

The present experiment was designed to study the relative roles of stimulus patterns and response patterns in determining recency effects. In the four stimulus–two response design introduced by Bertelson (1965) and used by Smith (1968) and Peeke and Stone (1972), each RT except that to the initial stimulus was categorized into one of three mutually exclusive categories: identical—a stimulus repetition implying a response repetition; equivalent—a stimulus alternation resulting in a response repetition; or different—a stimulus alternation resulting in a response alternation. In such a choice RT paradigm, *S*s react with a particular response to each stimulus so that the response cannot alternate when the stimulus repeats. The present experiment introduced a task which involved *S*s predicting which of two stimuli would occur on a given trial and then responding to the stimulus by making one response if they had been correct or an alternative response if they had been incorrect. With such a design, the alternation or repetition of a response was independent of the stimulus patterns and therefore the four categories, stimulus repetition–response repetition, stimulus repetition–response alternation, stimulus alternation–response repetition, and stimulus alternation–response alternation, were possible.

Method.—The 20 *S*s were volunteers from introductory psychology classes at Virginia Polytechnic Institute and State University. A 1.5 × 3 cm. digital readout was mounted in a partition which was located on a table and separated *E* from *S*. Prior to each of 200 choice RT trials, *S* verbally predicted which stimulus would appear and pressed a corresponding prediction switch. Following a prediction, *S*s grasped a right- and a left-hand handle, each handle encasing a sensitive microswitch. Ten *S*s pressed the right switch when the prediction outcome was correct and the left switch when the prediction outcome was incorrect; the remaining 10 *S*s used the reverse relationship (i.e., left following correct and right following incorrect prediction outcomes). The choice response stopped a Hunter Klockcounter. As determined by a punched paper tape, 50% of *S*s' predictions were correct, and the number of response alternations equaled the number of response repetitions.

TABLE 1
MEAN LATENCIES AND STANDARD DEVIATIONS FOR THE EIGHT OUTCOMES CATEGORIES DETERMINED BY TWO PREDICTION OUTCOMES AND FOUR STIMULUS–RESPONSE PATTERNS

Second-order pattern	Prediction outcome	Mean RT	*SD*
Stimulus repetition–response repetition	Correct	547	63.7
	Incorrect	689	95.5
Stimulus repetition–response alternation	Correct	597	96.8
	Incorrect	708	107.6
Stimulus alternation–response repetition	Correct	610	77.6
	Incorrect	729	113.0
Stimulus alternation–response alternation	Correct	594	83.6
	Incorrect	638	90.

Note.—Mean RTs and *SD*s are given in milliseconds.

---

[1] This research was supported partially by the Small Grants Program of the College of Arts and Sciences, Virginia Polytechnic Institute and State University.

[2] Requests for reprints should be sent to E. Scott Geller, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

Results and discussion. Latencies for anticipatory and inaccurate choices were excluded from the analysis and were fewer than 5% for each *S*. The analysis involved categorizing Trials 2–200 according to prediction outcome and then into the categories determined by the combination of the second-order stimulus and response patterns, i.e., stimulus repetition–response repetition, stimulus repetition–response alternation, stimulus alternation–response repetition, and stimulus alternation–response alternation.

The means and standard deviations of *S*s' mean latencies for the eight categories are presented in Table 1. A salient observation was that regardless of the prediction outcome choice RT was shorter in categories stimulus repetition–response repetition and stimulus alternation–response alternation than in stimulus alternation–response repetition and stimulus repetition–response alternation, respectively. An overall analysis of variance, a 2 (Prediction Outcome) × 2 (Stimulus: Repetition vs. Alternation) × 2 (Response: Repetition vs. Alternation) factorial, indicated a main effect of prediction outcome, $F\ (1, 19) = 85.46$, but no effect of stimulus or response, both $ps > .10$. The three second-order interactions were significant: Prediction Outcome × Stimulus, $F\ (1, 19) = 14.07$, $p < .005$; Prediction Outcome × Response, $F\ (1, 19) = 12.73$, $p < .005$; and Stimulus × Response, $F\ (1, 19) = 31.64$, $p < .001$. The third-order interaction did not reach significance at the .01 level, $F\ (1, 19) = 4.42$.

Apparently the second-order interactions involving prediction outcome were due to faster reactions to correctly predicted stimulus repetitions than to correctly predicted stimulus alternations, and faster reactions to incorrectly predicted stimulus alternations than to incorrectly predicted stimulus repetitions. That is, for either prediction outcome, *S*s were faster when they predicted a repetition than when they predicted an alternation.

Although *S*s' choice latencies of the present study appear to be longer than those of *S*s required to identify the stimulus,[3] the results of the present study suggest that the effects of recency and prediction outcome on choice RT are similar for the two paradigms. The observed effect of prediction outcome in the discrete RT design has been well established (e.g., Whitman & Geller, 1972). Furthermore, the progressively slower reactions to the categories stimulus repetition–response repetition, stimulus alternation–response alternation, and stimulus alternation–response repetition found with the present task have been observed in traditional choice RT experiments (Peeke & Stone, 1972; Smith, 1968).

Peeke and Stone (1972) suggested that choice RT was longest following stimulus alternation–response repetition because *S*s hesitate to make the same response to more than one stimulus. The present results indicated that *S*s also hesitate to make a different response to the same stimulus. In particular, *S*s responded faster when the repetition or alternation of the response corresponded to the repetition or alternation of the stimulus. One possible interpretation of this result was that *S*s were slower to repeat a response to alternated stimuli or to alternate responses to repeated stimuli because our technological environment rarely demands a given response to different signals or different responses to the same signal.

---

[3] Whitman, C. P. & Geller, E. S. Unpublished data (with the same apparatus). Virginia Polytechnic Institute and State University, 1972.

REFERENCES

Bertelson, P. Serial choice reaction time as a function of response versus signal-and-response repetition. *Nature*, 1965, **206**, 217–218.

Hale, D. J. Sequential effects in a two-choice serial reaction task. *Quarterly Journal of Experimental Psychology*, 1967, **14**, 113–114.

Kornblum, S. Sequential determinants of information processing in serial and discrete reaction time. *Psychological Review*, 1969, **76**, 113–131.

Peeke, S. C., & Stone, G. C. Sequential effects in two- and four-choice tasks. *Journal of Experimental Psychology*, 1972, **92**, 111–116.

Remington, R. J. Analysis of sequential effects in choice reaction times. *Journal of Experimental Psychology*, 1969, **82**, 250–257.

Smith, M. C. Repetition effect and short term memory. *Journal of Experimental Psychology*, 1968, **77**, 435–439.

Whitman, C. P., & Geller, E. S. Sequential effects of stimulus probability and prediction outcome on choice reaction time. *Journal of Experimental Psychology*, 1972, **93**, 373–378.

(Received February 18, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 468–470

IMAGERY CODES AND HUMAN INFORMATION RETRIEVAL

JOHN G. SEAMON[1]
New York University

In a short-term recognition memory task, *S*s who subvocally rehearsed or generated separate visual images for word stimuli showed a linear increase in response time with an increase in stimuli, while *S*s who generated an interactive imaginary scene showed no increase in latency. These data suggest a high-speed serial exhaustive retrieval process for the subvocal rehearsal and separate imagery conditions and a parallel process with constant comparison time for the interactive imagery condition. This implies that retrieval processes are dependent upon encoding strategies.

What are the cognitive processes involved in retrieving symbolic information from short-term memory (STM) and what do they imply about the operation of the memory system? Sternberg's (1966) study of short-term recognition memory has indicated that presence information is retrieved by a high-speed serial exhaustive scan of the contents of immediate memory. In that study, *S*s were presented with a short string of digits followed by a probe digit a few seconds later. The *S*s were required to indicate whether the probe was a member of the preceding set by making a YES or a NO response. Reaction time (RT) was found to increase linearly when set size was varied from one to six digits, with no difference in the slopes for the positive and negative response functions. These data suggest that the probe was serially compared to each item in the memory set before a response decision was made. Changing to a recall task, Sternberg (1967) found evidence for a serial self-terminating scan when *S*s had to recall the item which followed the probe position in a list of serial items.

One may wonder if a serial comparison process is a general characteristic of STM retrieval or the result of a particular encoding strategy employed by *S*s at the time of input. The present experiment sought to test this possibility by varying encoding conditions to determine their effect upon retrieval. Bower (1970) varied imagery strategies in a free-recall task and found significantly better recall with instructions to relate two word sets into single interactive images than instructions to generate two separate images for each set. Use of both imagery conditions in the STM-recognition paradigm, in addition to the traditional silent rehearsal condition, would appear to provide alternative encoding operations with which to determine if retrieval processes are dependent upon encoding strategies. Based on research with verbal and pictorial stimuli (Sternberg, 1969), a serial retrieval process with a scan rate of 18 to 26 items per second may be expected for the silent rehearsal and separate image conditions. Use of an interactive image, however, would appear to unify temporally and spatially the information in memory and may thus permit a parallel retrieval process in which each item is examined simultaneously.

Method.—Eighteen students from New York University served as *S*s to fulfill a course requirement. Three groups of 6 *S*s were compared, with memory set size and instructions varied in a Latin-square design, with three orders of set size and 2 *S*s per order for each instructional condition.

The *S*s saw one, two, or three simple English nouns on a rear-projection screen followed 2.5 sec. later by a short auditory warning signal and a probe word. The stimuli for a single trial were displayed simultaneously, with the display time calculated by multiplying the number of words shown by 5 sec. A total of 36 trials was presented for each set size in blocked format with 3 sec. between each trial. The first 12 trials in each block were considered practice. New words which varied in length from three to nine letters were used on every trial, with half of the responses positive and half negative. For positive responses, all serial positions were equally probable. The *S*s in all instructional conditions were encouraged to respond quickly and accurately by pressing either a YES button with their left index finger or a NO button with their right.

The first group of *S*s was instructed to rehearse the stimuli subvocally during their presentation and the blank period preceding the probe. This is procedurally similar to that employed in previous experiments. The second group of *S*s was instructed to generate an image for each of the words shown for study and to keep the images spatially separated and noninteracting at all times. These *S*s were told to hold the images in immediate memory by thinking about each in succession until the probe was presented and to respond YES only if the probe word represented one of the images generated. If *S*s were unable to form an image for any of the stimuli, or unable to hold them until the probe was presented, they were told to respond verbally "Can't do it" instead of making a YES or NO response. The last group of *S*s was also instructed to form an image of each of the study words, with an additional requirement of putting the images together into a single interactive scene. These *S*s were told to hold the imaginary scene by concentrating on it until the probe word was presented and to respond YES only

---

[1] Requests for reprints should be sent to John G. Seamon, who is now at the Department of Psychology, Wesleyan University, Middletown, Connecticut 06457.

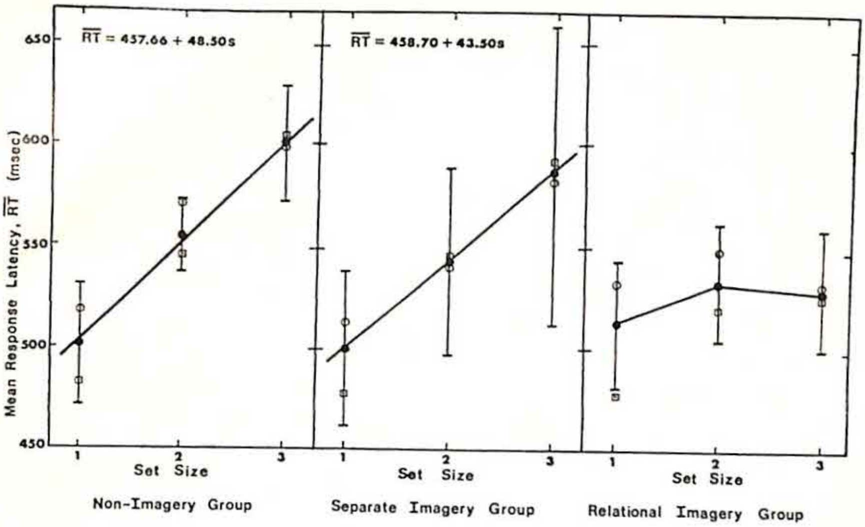


FIG. 1. The relationship between RT and memory set size for each of the three experimental groups. (Mean latency for positive [open squares] and negative [open circles] responses are based on the individual *S* means. Overall means for each set size based on the individual *S* means for all responses [filled circles] are shown with ± 1σ. Least-squares linear functions were applied to the overall means for the Nonimagery and Separate Imagery groups.)

if the probe represented one of the objects pictured in the scene. As in the previous group, *S*s were asked to respond verbally "Can't do it" if they could not comply with the instructions. Thus, using the three word set of VOLCANO–KNAPSACK–GORILLA as an example, *S*s in the Nonimagery group simply rehearsed the three words, *S*s in the Separate Imagery group formed a separate visual image for each of the three words, and *S*s in the Relational Imagery group might have imagined a gorilla wearing a knapsack jumping into a volcano. Each group saw the same stimuli and probes, with the order of presentation of each set size balanced within each group and equivalent between groups. Appropriate instructional examples were provided for each group to insure that all *S*s understood their task.

Results and discussion.—Errors for the Nonimagery, Separate Imagery, and Relational Imagery groups occurred on 4.16%, 2.08%, and 1.38% of the trials, respectively, and are not included in the analysis. Responses of "Can't do it" were similarly infrequent. They occurred on 1.38% of the trials for the Separate Imagery group and .69% of the trials for the Relational Imagery group. Mean latencies for correct responses are shown in Fig. 1. An increase in the number of items held in memory produced a linear increase in RT for the Nonimagery and Separate Imagery groups. Memory set size, however, had no effect on RT for the Relational Imagery group. Instructional set clearly resulted in differential effects on the response time variable. An analysis of variance for RT over memory set size found a significant effect of set size for the Nonimagery group, F (2, 8) = 6.35, $p < .05$, while nonsignificant results were obtained for the Separate Imagery group, F (3, 8) = 2.83, and the Relational Imagery group, F (2, 8) = 1.83, both $ps > .05$. The failure of the Separate Imagery group to reach significance may be attributed to the high degree of variability due primarily to two exceptional *S*s in this condition, one very fast and the other very slow. A trend analysis revealed a significant linear trend for the Nonimagery group, F (1, 8) = 12.66, $p < .01$, and the Separate Imagery group, F (1, 8) = 5.66, $p < .05$, with the quadratic component for both conditions less than 1. A nonsignificant linear trend, F (1, 8) = 2.23, and quadratic trend, F (1, 8) = 1.52, both $ps > .05$, were found for the Relational Imagery condition. A comparison of the Relational Imagery group over all set sizes with the Nonimagery and Separate Imagery groups at Set Size 1 shows no significant difference in RT ($F < 1.0$), indicating that *S*s using the relational imagery strategy took no longer to respond than *S*s using other strategies at Set Size 1.

Linear regression accounts for 99.70% and 99.98% of the variance for the Nonimagery and Separate Imagery groups. The linear functions for these groups are very similar, with the slope estimates within the range of 38 to 56 msec. reported by Sternberg (1969) with digits, shapes, and faces as stimuli. The linear increase in RT suggests a serial retrieval process for both groups. Excluding Set Size 1, which typically shows positives to be faster than negatives, the slopes for the positive and negative response functions differ by 4 msec. for the Separate Imagery group and 30 msec. for the Nonimagery group. In both cases, the slope for the positive function is greater than the negative function. The data are inconsistent with a self-terminating scan in which *S*s stop their search when a match has been found. This would predict the slope for positives to be half that of negatives as, on the average, only half of the items have to be scanned before a match is found for positives, while all must be examined for negatives. Since the slope estimate for negative responses is not larger than that for positives in either group, this argues for a high-speed serial exhaustive comparison process whose rate, based on the slopes for all responses, is between 20 and 23 items per second. Instructions to generate an image for each of the stimuli did not alter the underlying retrieval process.

A qualitatively different retrieval model appears necessary to account for the data of the Relational Imagery group. Since RT did not increase with set size, a serial retrieval process in which the probe is successively compared to the items in memory may be rejected. Instead, the data are consistent with a parallel process in which the probe is simultaneously compared to all items in the memory set with each comparison taking a constant amount of time. Thus three items may be examined as quickly as two or one. This indicates that recent theorizing by Paivio (1969) and Bower (1970), which suggests that the verbal system is primarily a serial processing system and the visual imagery system is a parallel processing system, is an oversimplification. Rather, the visual imagery system can be either serial or parallel depending upon encoding conditions. It remains to be seen if the same is true of the verbal system.

The results suggest that retrieval from STM should not be characterized as a simple serial process even within the limits defined by Sternberg (1966). The data strongly imply that retrieval processes are dependent upon encoding strategies. In accord with the hypothesis, if information is temporally or spatially separated in memory at the time of input, retrieval appears to require a serial search regardless of whether the verbal or visual imagery system is used. However, if the information is unitized, as in the present experiment by relational imagery instructions, retrieval may be described by a parallel model. A general theory of retrieval would thus seem to necessitate both serial and parallel processes. Additional research will have to determine the limit to the amount of information that can be unitized and retrieved in parallel by use of the relational imagery strategy. One suspects that the limit should not exceed Miller's (1956) STM capacity parameter of seven plus or minus two.

REFERENCES

BOWER, G. H. Analysis of a mnemonic device. *American Scientist*, 1970, **58**, 496–510.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **53**, 81–97.

PAIVIO, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–263.

STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

STERNBERG, S. Retrieval of contextual information from memory. *Psychonomic Science*, 1967, **8**, 55–56.

STERNBERG, S. Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457.

(Received March 13, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 471–472

PRESENTATION OF CORRECT AND INCORRECT ITEMS IN VERBAL DISCRIMINATION LEARNING

ROBERT W. NEWBY[1] AND ROBERT K. YOUNG

University of Texas at Austin

Three groups of *Ss* each learned a verbal discrimination (VD) list for six study–test trials. On the study trials, only correct responses were shown to the C group, only incorrect responses were shown to the I group, and correct and incorrect responses were shown to the CI group. Test trials were identical for all three groups. The results indicated that the performance of the C group was higher than that of the CI group, while the performance of the I group was lower. The results also indicated that the performance of the C and CI groups tended to improve over trials, while the performance of the I group showed an initial drop followed by a gradual increase over trials. The results were explained in terms of the frequency theory of VD learning.

A recent study of Ingison and Ekstrand (1970) using the study–test procedure with verbal discrimination (VD) learning found that when *Ss* are presented only with the correct member of a pair on study trials, performance is significantly better than when both the correct and incorrect members are presented. According to Ingison and Ekstrand, presentation of only the correct items on study trials increases the subjective frequency difference between the correct and incorrect items in a VD pair compared to the same items learned under the standard study–test procedure. That is, extra frequency units accrue to the incorrect items on study trials during the standard study–trial procedure. Since subjective frequency difference is identified as the major cue in VD learning, Ingison and Ekstrand's results fit easily within the framework of the frequency theory of VD learning (Ekstrand, Wallace, & Underwood, 1966).

The purpose of the present paper is to extend the Ingison and Ekstrand (1970) study by the addition of a group which receives only incorrect items on study trials. According to frequency theory, this manipulation should initially facilitate learning. The reason for this is that after the first study trial all incorrect items would have a frequency of one because of the representational response, while the corresponding correct items would have no frequency units. Therefore, on the first test trial (after the first study trial), an instructed *S* can use Rule 2 of VD learning: Always choose the least frequent item, and performance should be facilitated. However, during this first test trial the correct item would obtain frequency units because it is perceived (representational response), pronounced (pronunciation response), and rehearsed (rehearsal of the correct alternative response). Therefore, by the second test trial (after the second observation trial), the subjective frequency difference should be near zero so that the task would have been made more difficult. According to the theory, frequency units subsequently accrue to the correct items at a faster rate than to the incorrect items so that the frequency difference from that point on would gradually increase and performance would improve.

In the present experiment, *Ss* learned a VD list by the study–test method. Three study-trial methods were used: Only correct items we represented, only incorrect items, or both items were presented with the correct item underlined. In general, the theory predicts that presentation of only correct items should lead to better performance than with presentation of both correct and incorrect, while presentation of only incorrect items should lead to poorer performance.

Method.—Each of 120 *Ss* received one of three types of study-trial methods: correct only (C), incorrect only (I), or the standard study–test procedure (CI). Two samples of words were used, and each member of a given pair within a sample served as the correct item in one list and the incorrect item in another list. Presentation method, sample, and item function (i.e., C vs. I) were combined factorially. In addition, *Ss* were run in groups of five, two groups to a given list. All *Ss* were given six trials on the VD task, and the design can be summarized as a 3 × 2 × 2 × 2 × 6 mixed factorial ($n = 5$ per cell).

Two samples of 30 adjectives each were selected from Melton and Safier's (Hilgard, 1951) list and were used to construct two 15-pair VD lists. Interpair and intrapair associations were held as low as possible by visual examination of the lists.

Each list was presented by means of a Kodak 650 Carousel projector with an attached automatic timer. The lists were presented at a 3-sec. rate with a 12-sec. intertrial interval. Each of the six test trials was preceded by a study trial, and four different orders of study and test trials were used to control for possible serial learning effects. Study trials for the C group consisted of the presentation of each correct item at a 3-sec. rate; for the I group, the same procedure was used except that the incorrect items were presented; for the CI group, both items were presented with the correct item underlined. Within a given sample, test trials were identical for all groups. All *Ss* were given standard

---

[1] Requests for reprints should be sent to Robert W. Newby, Department of Psychology, University of Texas at Austin, Austin, Texas 78712.

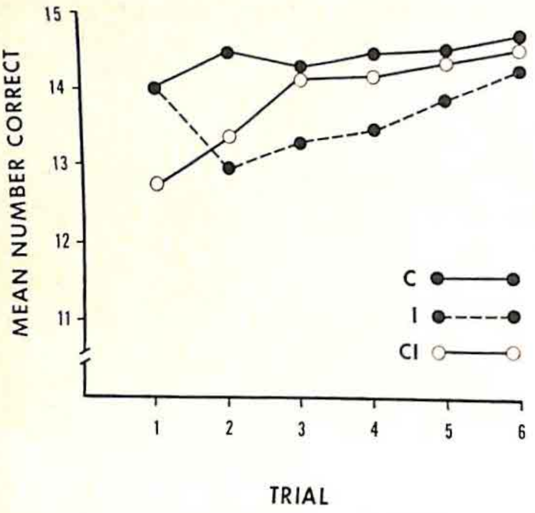


FIG. 1. Mean correct responses per trial for the C, I, and CI groups.

VD instructions and, in addition, *S*s in the C and I groups were told they were being presented correct or incorrect items, respectively. The *S*s were taken from introductory psychology classes and were naive to human learning experiments. No *S*s were dropped from the experiment.

Results and discussion.—The analysis of correct responses over the six trials of learning yielded a significant main effect for method of presentation, F (2, 96) = 5.43, $p < .01$, indicating that the information presented during the study trials influenced performance on the test trials. The trials main effect was also significant, F (5, 480) = 11.36, $p < .01$, and indicated that *S*s' performance improved over trials.

The Method of Presentation × Trials interaction was found to be significant, F (10, 480) = 6.15, $p < .01$, and is plotted in Fig. 1. As can be seen in Fig. 1, performance in the C group was consistently better than performance in the CI group. In addition, the I group was initially at the same performance level as the C group, but by the second trial the I group fell below the performance level of both the C and CI groups. These data support predictions made from frequency theory. Finally, word sample was found to be significant, F (1, 96) = 5.43, $p < .05$. Since this variable did not interact with other variables, it would appear to have relatively little behavioral significance. No other F reached significance.

Much of the Trials × Method of Presentation interaction occurs on the first two test trials. To describe these results further, the data of these two trials were analyzed alone. In this analysis, both method of presentation and the Method of Presentation × Trials interaction were identically significant, F (2, 96) = 9.09, $p < .001$. Both results are in agreement with the previous analysis. A third term, the Word Sample × Item Function × Trials interaction, was significant, F (1, 96) = 4.00, $p < .05$. However, considering the large number of F tests performed, this difference would seem to have been the result of sampling error. No other F was significant.

The performance of the C and CI groups replicates the results obtained by Ingison and Ekstrand (1970). These data support the notion that extra frequency units accrue to the incorrect items during the correct-plus-incorrect presentation on study trials, thereby decreasing the frequency difference between correct and incorrect items in accord with frequency theory.

The performance of the I group also supports predictions from frequency theory. Presentation of only the incorrect items on study trials should lead to slower learning than with either the correct-only or correct-plus-incorrect procedures. The data from the I group suggest that frequency units are added to both the correct and incorrect items on test trials as well as to the incorrect items on the study trials. This would help to explain the drop in performance in the I group from Trial 1 to Trial 2 and the subsequent improvement over the remaining trials. Had not frequency units accrued to the correct items on the test trials, *S*s could have continued to use Rule 2 and their performance would have paralleled that of the C group.

It is apparent from the present results that more occurs on a test trial than a simple measurement of performance. Specifically, learning of the correct response can occur only on test trials for the I group and, if the theory is correct, frequency accrues to both the C and I items during a test trial influencing subsequent learning.

Finally, the results of the present experiment are not intuitively obvious. That is, it would seem that in a two-choice situation such as VD learning, presentation of A or no presentation of A would provide *S* with identical information yielding essentially the same results in both cases. Such was not the case, and the theory was easily able to predict the results of the experiment.

REFERENCES

EKSTRAND, B. R., WALLACE, W. P., & UNDERWOOD, B. J. A frequency theory of verbal discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

INGISON, L. J., & EKSTRAND, B. R. Effects of study time, method of presentation, word frequency, and word abstractness on verbal discrimination learning. *Journal of Experimental Psychology*, 1970, **85**, 249–254.

(Received March 15, 1972)

Journal of Experimental Psychology
1972, Vol. 96, No. 2, 473–475

ACQUISITION AND FAMILIARIZATION HYPOTHESES IN PAIRED-ASSOCIATE LEARNING

ROBERT K. YOUNG,[1] ROBERT W. NEWBY, AND TERRY G. HAMON

University of Texas at Austin

If low-meaningful responses previously have been associated to high-imagery (or meaningful) stimuli rather than low-imagery stimuli, they are learned more easily in a subsequent free-recall task. The familiarization hypothesis states that differential familiarization accounts for this difference while the acquisition hypothesis states that the responses acquire some of the attributes of the stimuli during paired-associate (PA) learning. The present experiment tests these two hypotheses by employing the PA responses in a subsequent verbal discrimination task. In this paradigm, the acquisition hypothesis predicts positive transfer while the familiarization hypothesis predicts negative transfer. Positive transfer was obtained, and the data are interpreted as being supportive of the acquisition hypothesis.

A recent series of studies attempted to investigate the acquisition of such variables as meaningfulness (M) and imagery (I) in the laboratory. In general, the paradigm employed was one in which a paired-associate (PA) list was first learned followed by the learning of the PA responses in a free-recall list (Young & Barton, 1972). The M, I, and organization of the stimuli were varied during PA learning in various experiments. The results were consistent in that the higher the M, I, etc., on the stimulus side, the more rapid was the subsequent learning of the trigrams previously used as the PA responses. These results were taken as supporting the acquisition hypothesis: During PA learning a response takes on or acquires one or more of the attributes of the stimulus to which it is associated.

One of the nagging difficulties in the interpretation of the data of these experiments was the difficulty in ruling out differential familiarization as an alternative explanation. That is, the responses given correctly more often during PA learning may have been more familiar and, if so, should have been easier to learn during subsequent learning. This explanation does not require that differential I or M be acquired by the responses during PA learning. Instead, differential familiarization may be the variable influencing the results of this series of experiments.

Verbal discrimination (VD) learning may provide a method by which a choice can be made between these two alternate explanations. Initially the differential effects of I and familiarization upon VD learning should be noted. Paivio and Rowe (1970), for example, have demonstrated more rapid VD learning under high-I than under low-I conditions. In contrast Underwood and Freund (1970) have found that difficulty of VD learning increases as degree of familiarization, defined in terms of the number of times an item is exposed to *S* in the laboratory, increases.

Based on these results, if one of the attributes of the stimuli, such as I, were acquired by the low-M responses during PA learning, subsequent VD learning using the PA responses should be facilitated. But, if none of the attributes of the stimuli were acquired by the responses during PA learning, subsequent VD learning using the familiarized PA responses should be retarded. Thus the acquisition hypothesis predicts that VD learning of low-M responses previously associated to high-I stimuli will be facilitated while the familiarization hypothesis predicts VD learning will be retarded under exactly the same conditions.

Experiment I.—Each of 60 *S*s were given eight trials on a 16-pair PA list composed of noun stimuli and CVC responses and then were given five trials on a 16-pair VD list composed of CVC pairs. Half the CVC pairs in the VD list were constructed from the previously learned PA responses, while the other half of the pairs, the control pairs, were completely new and unfamiliar. The two sets of CVCs were counterbalanced such that for one half of the *S*s the experimental pairs in the VD list were made from CVC Set 1 and the control pairs were made from Set 2. This order was reversed for the other half of the *S*s. Two groups of PA stimuli were used and the noun and CVC sets were combined factorially to make four PA lists with noun stimuli and CVC responses. All *S*s learned the same VD list with differences in conditions being defined by the PA list learned. The design can be summarized as a $2 \times 2 \times 2 \times 5$ mixed factorial ($n = 15$ per cell). The verbal materials were taken from the same sources as were those in Young and Barton (1972). The nouns had A or AA Thorndike-Lorge ratings, and the CVC lists had median association values of 20 with a range of 17 to 25. Formal and meaningful similarity were held as low as possible.

Both the PA and VD lists were presented by a Kodak 650 Carousel projector with an attached automatic timer and all lists were learned by the study–test method at a 3-sec. rate with a 12-sec. intertrial interval. Four different orders of presentation for both the PA and VD lists were used to control for possible serial learning effects. Approxi-

---

[1] Requests for reprints should be sent to Robert K. Young, Department of Psychology, University of Texas at Austin, Austin, Texas 78712.

U. S. POSTAL SERVICE

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

(Act of August 12, 1970: Section 3685. Title 39. United States Code)

SEE INSTRUCTIONS ON PAGE 2 (REVERSE)

1. TITLE OF PUBLICATION JOURNAL OF EXPERIMENTAL PSYCHOLOGY	2. DATE OF FILING 9/27/72

3. FREQUENCY OF ISSUE
MONTHLY

4. LOCATION OF KNOWN OFFICE OF PUBLICATION *(Street, city, county, state, ZIP code) (Not printers)*
Prince & Lemon Sts., Lancaster, Lancaster, Pa. 17604

5. LOCATION OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS *(Not printers)*
1200 17th Street, N. W. Washington, D. C. 20036

6. NAMES AND ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR

PUBLISHER *(Name and address)*
American Psychological Association, 1200 17th St., N.W., Washington, D. E. 20036

EDITOR *(Name and address)*
David A Grant, University of Wisconsin, Madison, Wisconsin

MANAGER EDITOR *(Name and address)*
Harold P. Van Cott, 1200 17th St., N.W., Washington, D. C. 20036

7. OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given.)*

NAME	ADDRESS
American Psychological Association	1200 17th St., N.W., Washington, D.C. 20036

8. KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none, so state)*

NAME	ADDRESS
None	

9. FOR OPTIONAL COMPLETION BY PUBLISHERS MAILING AT THE REGULAR RATES (Section 132.121, Postal Service Manual)

39 U. S. C. 3626 provides in pertinent part: "No person who would have been entitled to mail matter under former section 4359 of this title shall mail such matter at the rates provided under this subsection unless he files annually with the Postal Service a written request for permission to mail matter at such rates."

In accordance with the provisions of this statute, I hereby request permission to mail the publication named in Item 1 at the reduced postage rates presently authorized by 39 U. S. C. 3626.

(Signature and title of editor, publisher, business manager, or owner)

10. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 132.122, Postal Manual)*

(Check one)

The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes	☒ Have not changed during preceding 12 months	☐ Have changed during preceding 12 months	*(If changed, publisher must submit explanation of change with this statement.)*

11. EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NUMBER OF COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A. TOTAL NO. COPIES PRINTED *(Net Press Run)*	5,221	5,262
B. PAID CIRCULATION 1. SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	None	None
2. MAIL SUBSCRIPTIONS	3,997	3,952
C. TOTAL PAID CIRCULATION	3,997	3,952
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS 1. SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	150	155
2. COPIES DISTRIBUTED TO NEWS AGENTS, BUT NOT SOLD	None	None
E. TOTAL DISTRIBUTION *(Sum of C and D)*	4,147	4,107
F. OFFICE USE, LEFT-OVER, UNACCOUNTED, SPOILED AFTER PRINTING	1,074	1,155
G. TOTAL *(Sum of E & F–should equal net press run shown in A)*	5,221	5,262

I certify that the statements made by me above are correct and complete ▶ *(Signature of editor, publisher, business manager, or owner)*
Henry S. Salmon Jr Bus Mgr

PS Form 3526 July 1971

"Why should I buy Gerbrands relay programming equipment when I have been using other electro-mechanical systems for so many years?"

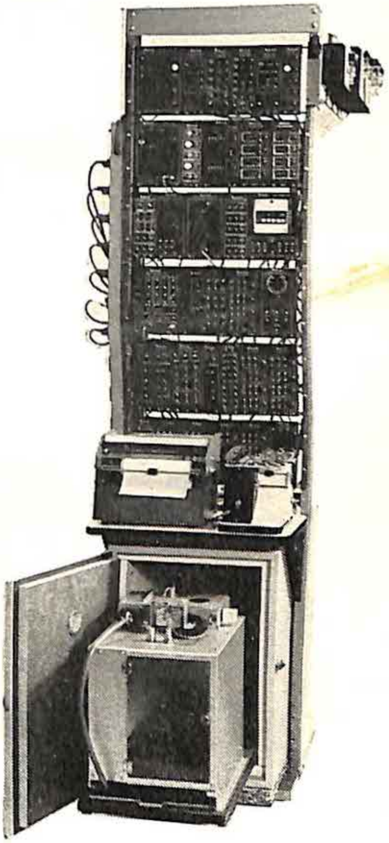
Why? Because most electromechanical programming systems presently in use were originally designed many years ago, and with most manufacturers concentrating on solid state and computer systems, few innovations have been added to relay programming equipment.

Until now.

Our relay programming line is completely new. It is easy to use (no engineering symbols or jargon), it is reliable, and it is innovative. Innovative with such new items as a Digital Interval Timer with push button control, a combination one per second/ten per second Pulse Generator with solid state switching, a Cam Stepper that incorporates two different mixed sequence steppers as well as an alternating stepper, and a rack mount Masking Noise Generator that sells for the low price of $75.00 fob Arlington, Massachusetts.

In addition, ours is an expanding line. As new technical developments become available we intend to incorporate these advances in our relay programming line wherever possible, in order to provide you with new and innovative instruments designed to make your electromechanical programming system more versatile and easier to program.

If you are thinking of purchasing a new electromechanical programming system or are looking for ways in which you can improve and expand your present system, write or phone for our latest catalog.



Ralph Gerbrands Company

8 BECK ROAD, ARLINGTON, MASSACHUSETTS 02174 (617) 648-6415

Bind Your Journal Issues Into Valuable Vigilant Volumes

Pertinent information you sometimes desperately need is too often in the Journal issue you cannot find. Single copies have a way of getting lost, misplaced or destroyed. It is better for you to let us permanently bind each journal into semi-annual or annual volumes—then your reference source is always complete, organized and instantly at your service. We call them "Vigilant Volumes"; they so carefully store and provide on instant notice so much knowledge of timely value.

The cost per "Vigilant Volume," by official commission from the journal publishers is but $5.95 per volume, permanently hard bound in washable buckram, gold embossed with period dates, journal name and special insignia—plus your name stamped in gold leaf. "Vigilant Volumes" are handsome library additions too, real conversation pieces. These volumes are bound in the authorized colors.

HOW TO ORDER

Simply ship your journal issues to us via parcel post, together with your name, address, AND REMITTANCE, at $5.95 per volume. Within 45 days after receipt, bound volumes will be shipped to you. Price subject to change without notice. Please remit 50¢ per volume for return transportation and handling.

PUBLISHER'S AUTHORIZED BINDERY SERVICE, Ltd.

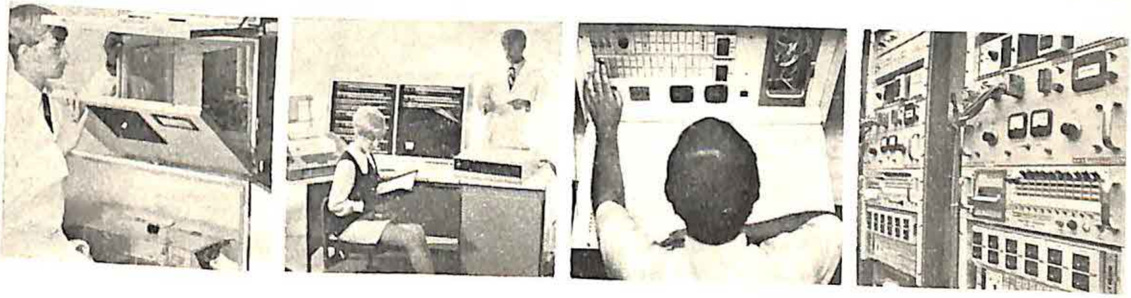
(Authorized Binders of All Journals)
4440 W. Roosevelt Rd.
Chicago, Illinois 60624

THE BEST
BEHAVIORAL RESEARCH
EQUIPMENT IS MADE
BY

BRS/LVE

Formerly

BRS-FORINGER
LEHIGH VALLEY ELECTRONICS



Write for our NEW Combined Catalog.

BRS/LVE
DIVISION of TECH SERV inc.

5301 Holland Drive • Beltsville, Md. 20705
Telephone 301/474-3400 • 800/638-0858

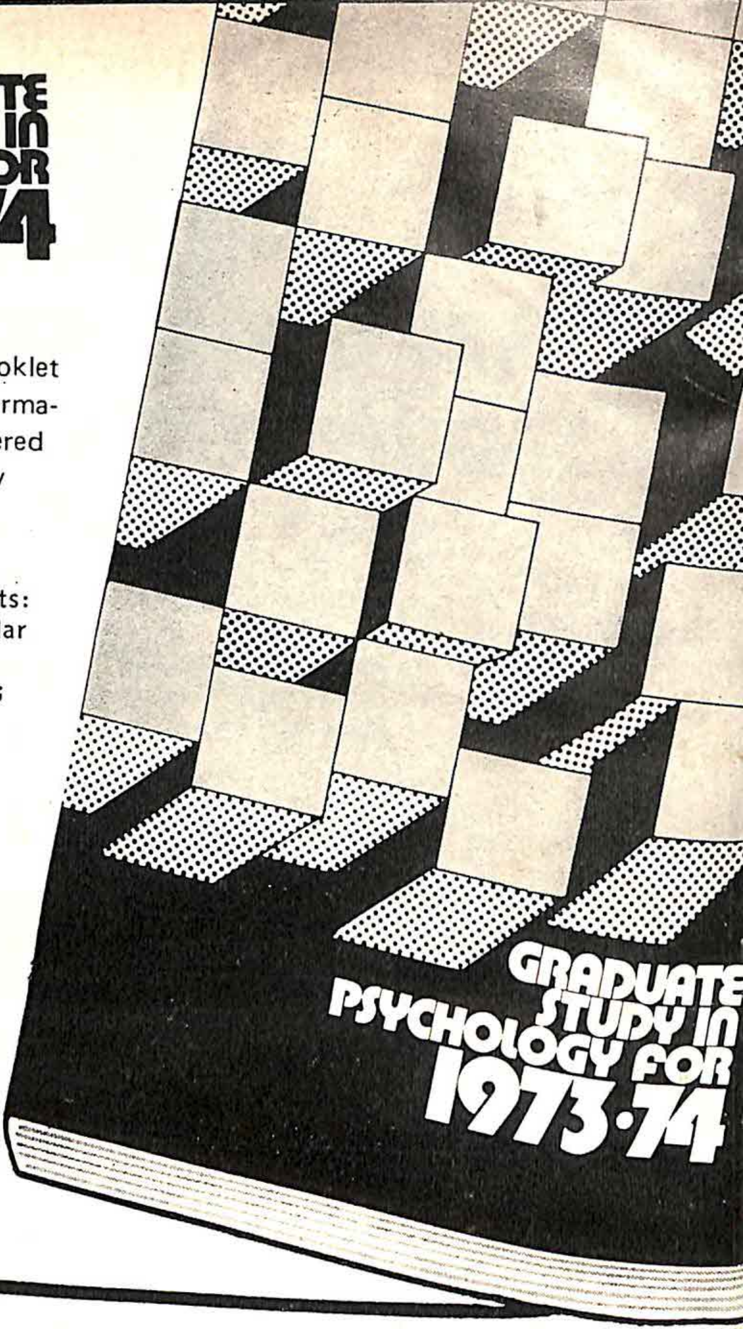
GRADUATE STUDY IN PSYCHOLOGY FOR 1973·74

The new sixth edition of this booklet provides relevant up-to-date information on 395 separately administered graduate programs in psychology in 315 universities and colleges.

Areas covered include enrollments: application for fellowships, scholar ships, or assistantships; types of assistantships and hours of work; government stipends; and post-doctoral arrangements. Current information has been supplied by the department chairman, his or her representative, or the program director.

$2 per copy

All orders amounting to $15 or less must be prepaid.


GRADUATE STUDY IN PSYCHOLOGY FOR 1973·74

☐ total enclosed for

☐ copies of **GRADUATE STUDY IN PSYCHOLOGY FOR 1973·74**

Order Department
AMERICAN PSYCHOLOGICAL ASSOCIATION
1200 Seventeenth Street, N.W.
Washington, D. C. 20036

Name ____________________

Address ____________________

____________________ Zip ____________